DOCUMENT RESUME

ED 187 739 TM 800 235

AUTHOR	Haney, Walt
TITLE	Summary of the Spring 1978 Conference of the National Consortium on Testing: June 5, 1978.
INSTITUTION	Huron Inst., Cambridge, Mass.
PUB DATE	5 Jun 78
NOTE	113p.; For related documents, see TM 800 236-237.
AVAILABLE FROM	Huron Institute, National Consortium on Testing, 123 Mt. Auburn Street, Cambridge, MA 02138 ($5.00)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Accountability; Court Litigation; *Educational Testing; *Minimum Competency Testing; Professional Associations; Public Opinion; *Standards; Testing Problems; Test Results
IDENTIFIERS	*National Consortium on Testing; Supreme Court

ABSTRACT

Remarks made at several panel discussions are summarized in this narrative report. The discussion topics and speakers include: (1) public education and testing--Tom Tomlinson, Ann Kahn, Herb Mack, and Jean Nazzaro, with remarks by Patricia Albjerg Graham; (2) standards regarding testing--Walt Haney, Barbara Lerner, Ann Cook, Willo White, and Bob Solomon; and (3) minimum competency testing--Edith Petrock, Peter Airasian, George Madaus, Merle McClung, and Hayes Mizzell. The appendices include a paper by Barbara Lerner on the Supreme Court and professional standards for testing; Willo White's memorandum, "Call for Information on Test Misuse and Abuse;" the Education Commission of the States' paper on policy issues concerning minimum competency testing; Peter Airasian's paper, "Pre-measurement Issues in Minimal Competency Testing;" and "Measurement Issues and Consequences Associated with Minimal Competency Testing," by George Madaus. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED187739

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SUMMARY OF THE SPRING 1978 CONFERENCE

OF THE

NATIONAL CONSORTIUM ON TESTING

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Paul L. Houts
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Prepared by

Walt Haney

Huron Institute
123 Mount Auburn Street
Cambridge, Mass. 02138

June 5, 1978

TM800 235

This summary has been prepared from notes and partial tape recordings of the substantive meetings of the conference. It does not represent a verbatim account of the conference.

TABLE OF CONTENTS

Introduction	page:
Vito Perrone	1
David K. Cohen	1
Panel on Public Education on Testing	
Tom Tomlinson	2
Ann Kahn	3
Herb Mack	4
Jean Nazzaro	5
Discussion	6
Remarks by Patricia A. Graham	
Patricia A. Graham	8
Discussion	10
Panel on Standards Regarding Testing	
Walt Haney	11
Barbara Lerner	13
Ann Cook	13
Willo White	17
Bob Solomon	18
Discussion	20
Panel on Minimum Competency Testing	
Edith Petrock	21
Peter Airasian	22
George Madaus	22
Merle McClung	23
Hayes Mizell	25
Discussion	27

Attachments

Barbara Lerner, "The Supreme Court and the APA, AERA, NCME Test Standards: Past References and Future Possibilities," DRAFT.

Willo White, memorandum, "Call for Information on Test Misuse and Abuse," May 19, 1978.

ECS policy paper.

Peter Airasian, "Pre-measurement Issues in Minimal Competency Testing."

George Madaus, "Measurement Issues and Consequences Associated with Minimal Competency Testing."

The Spring 1978 Membership Conference of the National Consortium on Testing was held May 15 and 16 at the Holiday Inn in Arlington, Virginia. This paper summarizes the substantive meetings on Monday and on Tuesday morning. Minutes of the Consortium's business meeting on Tuesday afternoon will be circulated separately.

INTRODUCTION

The meeting was opened on Monday morning with a welcome by Consortium President Vito Perrone. He noted that despite the tradition of bad weather for Consortium meetings in Washington, the attendance at the conference was encouraging--more than 60 people. In light of the growing interest in the Consortium and the full agenda for the spring conference, he expressed optimism regarding the future of the group

David K. Cohen, Executive Director of the Consortium, reported briefly on activities of the Huron Institute in providing staff work for the Consortium. He related that Huron had developed a program plan of activities which had been unanimously approved by the Consortium's Board of Directors in March. Since then Huron has been working to implement those plans, focusing initially on the program area of truth-in-testing and to a lesser extent on the federal government's role with respect to testing. At the same time, however, David noted that Huron staff had also been distracted from the formal plan of activities by several opportunities to participate in special meetings on testing. In March, for example, David Cohen and Walt Haney attended the HEW conference on achievement testing and the basic skills, and prepared a summary of the conference for NCT members.[1]

---

[1] Letter of March 17, 1978 from Walt Haney to Consortium members.

In April they attended a meeting of the Educational Forum in Arlington, Virginia to present a paper by Walt Haney and George Madaus on minimum competency testing.[2] In addition, David reported that Huron had spent considerable time responding to requests for information both from Consortium members and from other individuals and groups interested in the Consortium and in testing issues. In the future, he noted Huron would proceed to implement plans for staff work on the Consortium.

Panel on Public Education on Testing

The panel on public education on testing was chaired by Vito Perrone. The first presentation was by Tom Tomlinson of the National Institute of Education (NIE). Tom prefaced his remarks by saying that he was by no means presenting NIE policy regarding testing. Instead, he was presenting strictly his own views on the topic. He continued with an anecdote about his own daughter who had achieved relatively poorly in mathematics during her school career. Test scores she received at the end of high school, Tom observed, really had not told anyone anything new about his daughter's educational performance, but instead had served simply to confirm what was already known. In light of that experience, he said that a host of questions need to be considered before determining what sort of public education on testing needs to be undertaken. Among the questions he offered were:

- What does the public know about testing?
- What does the public think they know about testing?
- Who is the public?
- What should they be told about testing?

---

[2] Walt Haney and George Madaus, "Making Sense of the Minimum Competency Testing Movement." Paper presented to the Educational Forum, Crystal City, Virginia, April 14, 1978. A revised and expanded version of this paper will be distributed as a Consortium circular.

- What about classroom tests?
- What do teachers know about testing and what do they need to be told?
- Should I tell the public about the costs of doing things differently?
- What might I tell the public about alternatives to testing?
- What do kids thinks about testing and what do they need to be told about it?

In closing, Tom noted that we seem to be getting very agnostic about testing and that we need to be careful to avoid communicating the impression that nothing at all can be measured very well. There is a danger, he suggested, that such agnosticism might carry over to contribute to a wider agnosticism about public education in general. What we perhaps need to do, he continued, is to go back to seek consensus on the goals of education and then go about the job of measuring the attainment of those goals in an appropriate way.

The second speaker on the first panel was Ann Kahn of the National Congress of Parents and Teachers. In opening, Ann noted that in thinking about testing she empathizes with the sentiments expressed in a pair of bumper stickers she had seen recently: "You will pay for your sins" and "If you have already paid, please disregard this message." She noted that in her contacts with the general public, she has found very little understanding of the sort of issues regarding testing that brought the NCT together. It's often from parents, she thinks, that pressures for more testing are coming because of parents' concern over the quality of education their children are receiving. Legislators seem to be responding to dissatisfaction from both parents and the public generally. Another angle on the concern over testing, she notes, is demography. In her district for example, more than 50% of the voters have no children in the public schools. Since they have no direct contact with the schools, they seem to see testing as a means for knowing what's happening in the schools. Teachers who

oppose tests, Ann continued, generally do so because they are a pain in the neck, not because of the reasons that brought the Consortium together. She noted that there is a tendency for people in the Consortium to talk to people like themselves. However, if the Consortium is to have any real impact we have to think in much broader terms. From her experience, Ann judges that there is no mass of troops out there waiting to join in the Consortium's efforts. So she feels that we have to try to develop materials beyond policy papers and position statements. Instead we need to express our concerns in materials that have meaning for the average individual who would not normally read trade publications. What we should do, Ann advised, is to publish our concerns in vivid terms in mass circulation periodicals like the Lady's Home Journal and Family Circle. Unless we develop public support, Ann continued, whatever else we do will not matter much at all. In closing, she noted too that none of our efforts will have much meaning unless real alternatives to standardized tests are available. The public will not accept the notion that we should do away with tests and instead simply observe children and say how they're doing on that basis. Hence, she concluded, systematic alternatives to standardized tests and broad public education effforts will be crucial to the success of the Consortium's endeavors.

The third participant in the public education panel was Herb Mack of the North Dakota Study Group on Evaluation (NDSG). Herb described the experience of the NDSG in preparing a small booklet "Reading Tests--Do they Help or Hurt Your Child?" He said that this booklet had been very popular and all available copies of the booklet had been distributed. As a result of that experience and because of the NDSG's interest in public education regarding testing, the group has prepared a revised version of that booklet entitled "Standardized

Reading Tests: Do They Help or Hurt Your Child."[3] Other booklets which are or will be available are: Standardized Reading Tests:

- What Does that Score Mean
- How to Prepare Your Child
- Their Impact on Curriculum: Do we Test What's Taught or Teach What's Tested
- What are the Alternatives

Herb reported that the NDSG hopes to distribute these booklets widely so that they reach a broad audience in the sort of fashion Ann suggested as necessary. Herb also mentioned that while several of the booklets will be critical of standardized testing, the booklet on preparing children to take tests is still necessary. Though much standardized testing serves no useful prupose and is often irrelevant to the school curriculum, it seems unlikely that it will soon be done away with. In light of this reality, this booklet will seek to help parents and teachers prepare children for taking standardized tests so that they can do well on them and not be penalized for their lack of familiarity with that form of assessment.

Jean Nazzarro of the Council for Exceptional Children (CEC), the fourth participant on the public education panel, described her experiences with CEC in developing materials about testing and in conducting workshops regarding testing with exceptional children. Most of those efforts she noted, have been addressed to how tests can be unfair when used with exceptional children, public policy regarding testing and comprehensive assessment with special education children, and procedures for nondiscriminatory testing. Also she briefly

---

[3] Copies of this booklet are available at 50¢ a single copy (or 40¢ for bulk orders of ten or more) from The North Dakota Study Group on Evaluation; Center for Teaching and Learning; University of North Dakota; Grand Forks, North Dakota 58202 -- or from Ann Cook and Deborah Meier; 670 West End Avenue; New York, New York 10025.

described two sets of materials regarding testing prepared by CEC which are available for use by other groups.[4] In Jean's view, special education is a step ahead of education generally in terms of promoting sensible assessment practice. She noted, for example, that P.L. 94-142, Education for All Handicapped Children, requires individualized educational plans (IEPs) for all handicapped children served under that law. Precisely that sort of individualized planning is, however, just what is needed for all children. Jean also suggested that Consortium members would all benefit from more information exchange, both in terms of information on testing and in terms of positions or recommendations advocated by different member groups of the Consortium.

Discussion following panel presentations was spirited. Ann Cook said that it may be okay to talk about quality education and goals of education, but that the trouble is that we've gotten ourselves in a position in which again and again people judge the quality of education in terms of test scores. What we need to do, Ann recommended, it to help get people into the habit of thinking for themselves on this issue. We've got to get people out of the habit of calling the weather bureau to find out what the weather is like, she continued, and instead get them to stick their head out the door and look for themselves. Instead of giving kids reading tests to find out whether or not

---

[4] The two sets of materials "How Can Tests be Unfair?" and "Protection in Evaluation procedures: Nondiscriminatory Tesing." The former is a boxed set of materials for conducting a workshop on nondiscriminatory testing. The latter is a manual bound in a loose-leaf notebook designed to serve as a source material for a CEC Institute. More information on these materials is available from the Council on Exceptional Children, 1920 Association Drive, Reston, Virginia 22091. Jean also mentioned a "Nondiscriminatory Factsheet" available from the ERIC Clearinghouse for Handicapped and Gifted Children, 1920 Association Drive, Reston, Virginia 22091.

they can read, we need to just ask them to read something. Ann Kahn commented that there is a tremendous amout of misinformation around concerning testing. She cited the example of one legislator who was upset because a school in his community was only at the 70th percentile. "In his days in school, 70% meant a "C" and that was not good enough for his liking." David Prasse expressed his concern over appropriate strategies for dealing with testing issues. There is a need not just for parents to ask intelligent questions about testing, he maintained, but also for educators to give intelligent answers." In his view, much confusion regarding testing may have arisen simply bacause so many issues have gotten bound up into discussions about testing practices. As far as he is concerned, there really is no single issue of testing, but a multitude of different ones. Lillian Weber voiced her concern over an apparent tendency for people to equate the influence of parents and teachers regarding testing. Teachers, Lillian noted, generally do no have much say at all about what sort of standardized tests are given in their classrooms. Moreover, there seems to be an enormous mistrust of teachers nowadays. As a possible antidote to this misunderstanding and mistrust, Lillian suggested that it might be useful to have people write their reminiscences regarding what schooling and testing was like in their own school years--in the past, which many people look upon as being a time of higher educational quality.

Bill Pierce expressed his concern that members of the Consortium as professionals not accede to the polling approach to public policy. He noted, for example, that just because a test of sexual preferences might just now be a popular method of deciding on retaining or dismissing teachers in several states, this does not at all mean that we should support such a practice.

Other discussion revolved around the role of competition in American society and education and the tension between the educational and child care

responsibilities of the educational institution. Several people voiced support for development of better alternatives to standardized testing and for sharing information on such alternatives. Several others stressed the importance of the issues discussed in the first panel session and the urgent need for materials describing testing issues in a straightforward and simple manner that can be understood by parents and teachers and the public at large.

Remarks by Patricia Albjerg Graham

After lunch on Monday, Patricia Graham, Director of NIE, addressed the membership conference. In her remarks she raised several issues which people who use testing procedures need to think about, First, she said, we will all have to continue our struggles to ask pertinent questions about what is it that tests test. For example, she suggested, she is not at all convinced that we know how to distinguish when tests measure what children learn in schools and when they measure what children learn elsewhere. She noted that this confusion may have been used to good advantage by school people--to take credit when good things seem to be happening even though schools may not deserve all the credit; and to deny blame when not such good things seem to be happening even though schools may deserve some of the blame. Second, she suggested that we need to be clearer about whether we want to measure students or whether we want to measure schools. Personally, she has more interest in measuring schools because in the past schools have not recognized the extent to which they have been responsible for student achievement. In the past, she continued, this issue has been clouded by the practice of encouraging low achieving students to drop out ot school or not to come at all. However, now that we have a clearer commitment to educating all children, we

have to come to grips more clearly with the need to test what students are acquiring in schools and also, of course, with the issue of what students should acquire through schooling. If we insist on trying to measure students instead of schools, Ms Graham continued, we will probably be engaged more in measuring broader social outcomes than in measuring the real consequences of schooling.

Ms Graham's third point was that our confusion about testing is heightened by the similarity in form but difference in function between aptitude and achievement tests. While aptitude testing comprised much of the early work in the field of psychometrics, she continued, the requirements of achievement testing are quite different than that of aptitude testing. Specifically, she maintained, if achievement testing is to be of real help in identifying areas where students are not achieving what we would like, we need to know how achievement tests are related to what we want schools to teach.

Ms Graham also summed up the four main themes that had emerged from the HEW Conference in March:

1. Testing is not the basic issue; rather it is the establishment of fundamental attainable goals in such areas as basic skills and literacy.
2. Educational accountability is necessary and inevitable, and achievement tests are necessary but imperfect tools.
3. There was overwhelming opposition to the idea of a national test of educational achievement.
4. There was a strong desire for federal support for distribution of information about tests and their limits and uses; support for training to promote sensible utilization of tests in the field and expanded research on testing.

In closing, Ms Graham noted that many of the questions that have been and are being raised about testing are in truth deeper questions about what we want our schools to teach, what we want our children to learn, and what we want our society to become.

Following her remarks, Ms Graham entertained questions from the floor. Herb Rudman inquired as to HEW plans for promoting public education regarding testing. Ms Graham responded that while she could not speak for HEW in general, she could report that one activity in this area planned by NIE is a series of regional conferences regarding testing to which a variety of people with diverse interests in testing would be invited. In a follow up remark, Herb expressed concern that such conferences not turn into negative discussion of why tests should not be used, but instead attempt to provide useful information on how tests can be sensibly used as well as information on their limits. Ms Graham responded by noting that the federal government is one of the biggest promoters of test use, for example, concerning allocation of ESAA funds; that more broadly the pros and cons of testing need to be weighed against alternative sources of assessment information, and that the task of disseminating information which would be useful in helping peole make better educational decisions was a difficult one but one that NIE would be working on. In light of Ms Graham's questions concerning whether tests measure in-school or out-of-school learning, and her observation that the federal government is a big promoter of test use, Fritz Mosher of the Carnegie Corporation questioned whether more caution might be warranted regarding federal promotion of test use. Ms Graham responded that such caution probably is warranted but that the practical problem is that there often are no readily available alternatives

to standardized tests waiting to be plucked off the shelf to be used in their stead. Barbara Lieb-Brilhart of the Speech Communication Association expressed appreciation for Ms Graham's broad definition of literacy since she feels that it is in areas of literacy beyond reading and writing that we have special problems with assessment, and asked what NIE"s commitments in that area might be. In response, Ms Graham mentioned that appointment of a new Associate Director of NIE for Teaching and Learning, Sylvia Scribner, most recently of the Rockefeller University Center on Cognition. Ms Graham noted that one of the reasons NIE hired her was because of her interest in a broader conception of literacy.

Panel on Standards Regarding Testing

After Pat Graham's remarks, David Cohen introduced the second conference panel on the topic of standards for development, validation and use of tests. The first panelist in this session was Walt Haney of the Huron Institute. He commented on several past efforts to develop standards regarding testing; specifically three versions of the APA Ethical Standards of Psychologists (from 1953, 1963, and 1977) and three versions of technical standards on testing sponsored jointly by the APA, the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME), published in 1954, 1966, and 1974. He noted that the portions of Ethical Standards regarding testing have shrunk in size over the three versions of those standards. Nevertheless, the 1977 version of the Ethical Standards gives more prominence to the rights of test takers to know the basis on which they are judged relative to the attention given to test security--an issue that was given much more prominence in the 1953 and 1963 versions of the Ethical Standards.

The technical standards, Walt reported, were quite similar in format in the 1954 and 1966 versions. The 1974 version of those standards was, however, substantially revised in that it devoted a separate section to standards for test use. After describing the basic changes in these six sets standards over the last quarter-century, Walt drew attention to five issues that emerged for him as especially important in his review of the standards. First, regarding validity, he noted that althrough the three versions of the technical standards all mentioned the four types of validity, criterion-related validity (that is concurrent and predictive) received a disproportionate amount of attention in the 1974 standards. Regarding reliability, he noted that there seems to be a trend away from using various correlation coefficients (concerning internal consistency, alternative forms and test-retest reliability) as a means of judging test reliability, and more toward components of variance and standard error as useful means of assessing test reliability. Walt's third observation was that the portion of the 1974 technical standards regarding test use does not very clearly distinguish between alternative uses of tests nor does it very clearly delineate the implications of alternative uses for specific standards. Fourth, Walt observed that there seems to be a gap between the 1977 Ethnical Standards' recognition of test-takers' rights to information regarding the basis on which they might be judged and the lack of attention given to this issue in the 1974 technical standards (which give relatively more attention to test security). Walt's fifth point was that there is evidence of a gap between any of the versions of test standards and testing practice. Specifically, he suggested that anyone who seeks to reform testing practice with such standards ought to give closer attention to the relationship between

standards and practice.[5]

The second panelist on Monday afternoon was Barbara Lerner of the National Academy of Sciences. Barbara's remarks were based on her review of Supreme Court references to test standards and to federal agency guidelines on testing over the past seven years. Her review suggests that the Court is beginning to rely more heavily on the Joint APA, AERA and NCME Standards as an authoritative source for objective answers to technical and psychometric questions. She concluded her remarks with five recommendations regarding how the joint standards might be revised.[6]

The third panelist on the standards panel was Ann Cook of the Community Resources Institute. In opening, Ann mentioned that last year, a group of parents, administrators and teachers in New York City took the city's Board of Education to court in an effort to stop the city from administering the city-wide reading and math tests. The reason the case was initiated was because it had become apparent that the security of tests had been compromised. Some children had actually been coached on specific items from the tests. The contention of the plaintiffs was that prior knowledge of the tests for some children would adversely affect the standing of other children in the city on those tests and therefore the testing should be stopped. The group that brought the suit was not really so concerned with whether a test should be secure or not. In fact, she mentioned, there are some problems with whether a norm-referenced test really can be totally secure since it must have been

---

[5] A written and expanded version of Walt's remarks will be circulated to Consortium members within the next few weeks.

[6] A copy of the draft paper on which Barbara Lerner's remarks were based is appended to this report. Requests for reprints of the paper should be addressed to: Barbara Lerner, National Academy of Sciences, 2101 Constitution Avenue, N. W., Washington, D. C. 20418.

administered somewhere in order to develop norms. The real issue of concern to the groups concerned, however, was that students who had been coached on specific items from the tests would have an unfair advantage over children who had not been coached.

In Ann's view this case served to highlight many of the important issues concerned with the use of norm-referenced tests in real-life day-to-day school settings. The purpose of such tests is really to allow group comparisons. They are not designed, as even test publishers acknowledge, to tell us about individual children. From such tests, we do not get specific information on how individual children perform on specific items. To the extent we've been able to find out why kids score as they do, she continued, we've often found that kids get the right answer for the wrong reasons or that they get wrong answers for extreneous reasons, for example because they were nervous about the test conditions. So the real issue in the court case was the use of norm-referenced tests with individual children.

So what are these uses, Ann asked. We've had a lot of difficulty pinning the Board of Education down as to what the uses of the city-wide tests actually are. As a result of the court case, the group did finally get a statement of specific uses. Ann read from a sworn statement of Mr. Charles Schonhaut, an official of the New York Public School System:

> The reading test that is administered to grades 2 through 9 in New York City serves several purposes. Among them are:
>
> a) to rank schools as required by Education Law §2590-j (5)(a)
> b) to organize classes for the next school year
> c) for promotional requirements for students
> d) to evaluate reimbursable programs, a federal and state requirement
> e) to determine an individual student's eligibility for Title I programs
> f) for generating computer tapes with pupil information that is used by the Office of Educational Performance and the Office of Educational Statistics

g) to make comparisons with previous years
h) for articulation with different levels of the school system, i.e., elementary to junior high school, junior high school to high school.[7]

Ann noted that the uses cited for the city-wide math tests were essentially the same as those for the reading tests.

So, Ann continued, it is now clear that a single score on such city-wide tests coming back to the schools on computer printout will affect educational decisions made about individual children. What Mr. Schonhaut's comments did not mention, Ann remarked, is that in New York City we have specific grade-equivalent scores required for students' admission into special schools--not a range, a percentile or a stanine, but a specific grade equivalent score. Also, Mr. Schonhaut's statement did not mention two other areas. First was use of test scores as they affect parents' assessment of their children's performance. Most inner-city parents, Ann suggested, see grade equivalent scores as the most important indicator of their children's educational success or failure. Second, is the use of tests in pupil assessment. Students' grade equivalent scores can be affected simply because of the way tests are normed. Even though kids make real progress, they can appear to fall back simply because of the way grade equivalent scores are calculated. And parent, teacher (and certainly not kids) do not understand this. Also individual children's scores, Ann continued, can affect classes and teachers children get assigned to--for example, grade equivalent scores can affect whether or not kids get into

---

[7] Sworn statement of Dr. Charles Schonhaut, Senior Assistant for Education and Performance and Accountability to the Chancellor of the New York City Public School System in the case of Katherine James et al against Board of Education of the City of New York; Supreme Court of the State of New York, County of New York. Index No. 5836/77 Affidavit in Opposition to Preliminary Injunction.

foreign language, art, or sports programs. Ann also suggested that the use of grade-equivalent test scores to determine eligibility of schools to receive funds has a disincentive attached to it. If a school's performance gets too high, it may not be eligible to continue in a program. This results, she recounted, in lots of juggling in testing--with some children tested in the fall and maybe a different group tested in the spring.

Moreover, Ann went on, tests affect curriculum. The material used in reading classes, for example, begins to look very much like tests. Instead of teaching kids to appreciate reading and to read, reading clinics use test-like materials to teach them to become profocient test-takers. So we're subtly teaching kids what reading is about--it's about reading tests.

This has come about, Ann suggested, in part because of the way tests are reported back to schools. Usually they are reported as total test scores in grade equivalent terms--a form which is totally useless in terms of helping a teacher know exactly where a student needs help.

Ann noted too that the initiation of specific grade-equivalent requirements for grade promotion is leading to even more questionable practices with respect to testing. In the fall a school can have substantial numbers of kids reading well below grade level. Then in the spring prior to grade promotion, large numbers of them have miraculously been brought up to grade-level. This happens, Ann suggested, not by suddenly teaching them to read, but by coaching kids on tests. By having kids take tests time after time for a period of three months, schools have found that their scores can be raised into the acceptable range. They are not necessarily better readers, but they are more proficient test-takers.

Finally, Ann suggested that she is puzzled by the lack of test-publishers response to this phenomenon. If larger numbers of kids are getting to be

better test-takers and more are scoring at grade-level, eventually publishers will have to renorm their tests. We should all be ready for this, Ann advised because when it happens kids' scores will appear suddenly to drop. We've seen that sort of thing happen in New York, she remarked, when scores were suddenly inflated or deflated simply because tests were renormed. As a result Ann suggested, we all have to look much more carefully at what happens to specific kids who get labeled early in school as poor readers as they progress through school.

The fourth participant on the standards panel was Willo White of the American Psychological Association. She began by noting that the APA is in the process of becoming a member of the National Consortium on Testing and she hopes that the APA will soon become an official member. She noted too that the APA has been in the forefront of the effort to develop professional standards. It was, and still is as far as Willo knows, the first and only organization to develop an empirically derived set of ethical standards. She also voiced a concern that Walt Haney's presentation may have left a slightly misleading impression regarding the latest version of the APA ethical standards. While the number of principles dealing with testing have diminished in the latest version of the Ethical Standards, she noted that a number of concerns previously treated in separate principles now are subsumed under the single principle on assessment procedures. She noted too that the issue of ensuring adherence to such standards is a big one. Unfortunately, she noted, the APA has little power to bring sanctions against members who violate standards. The strongest sanction now available is to drop an individual from APA membership. However, such a procedure must be handled through the APA ethics committee. And since this process is usually quite combersome and time-consuming, it does not seem to be an effective procedure to ensuring adherence to standards. Willo

also mentioned that the APA testing committee is interested in gathering a set of critical incidents concerning testing in order to develop a case-book. She solicited the help of Consortium members in collecting such incidents. In closing, Willo mentioned that the APA testing committee is very interested in the issue of due process for test-takers and would like to develop some kind of a bill of rights for test takers.[8]

The fifth participant on the standards panel was Bob Solomon of the Educational Testing Service. Bob began by noting that the development of the 1954 joint standards was a major step forward in the effort to articulate standards in the testing profession, and the expansion of the standards in 1966 and 1974 and the revision now underway are attempts to meet the challenge of accountability by the measurement profession. One limitation of those standards, however, is their orientation toward the test instrument itself, and the relative lack of attention to related activities provided by agencies such as ETS. Test publishers, Bob continued, obviously do much more than develop, administer and score tests. For example, they also engage in research and development, they store and retrieve data about individuals, develop publications and other interpretative materials and undertake educational programs designed to improve the understanding of test users. Testing agencies have recognized their responsibilities to the individuals and institutions they serve, and also to the larger society, but more remains to be done. In

[8] Attached to this summary is a copy of a memo from Willo soliciting information on test use and abuse.

the past few months, ETS has taken a step forward in this direction by developing a set of "Principles, Policies and Procedural Guidelines Regarding ETS Products and Services." The guidelines, Bob went on, represent an intensive effort by over 150 of ETS's staff over a two year period. They are designed to ensure that ETS products conform to uniformly high standards with respect to six areas of basic importance:

> 1) Accountability, the responsibilities of ETS to those affected by its activities; 2) Confidentiality of Data, the rights to and limitations on access to data collected by ETS; 3) Product Accuracy and Timeliness, the control of quality and performance according to scheduled commitments; 4) Research and Development; 5) Tests and Measurement; 6) Advisory and Instructional Activities.

Bob continued to read from the preface to the ETS guidelines. Since copies of those guidelines were distributed to people attending the conference, this summary will not repeat his remarks.[9]

In a final point, Bob observed that all of the efforts to improve the mechanisms of accountability, laudable though they may be, do not necessarily go to the heart of the matter. The inner core of the problem of defining the responsibility of testing agencies, has more to do with value choices and the public policies that flow from them, than it does with guidelines for monitoring and improving services. Tests are not used because measurement experts develop them. Tests are used because they are seen by policy makers as useful instruments for improving social and educational efforts. Public and institutional policies that entail the use of examinations have developed because these policies are seen as the means by which certain objectives are realized.

---

[9] Additional copies of the ETS "Principles, Policies and Procedural Guidelines Regarding ETS Products and Services," are available from Information Services Div., Educational Testing Service, Princeton, New Jersey 08541.

They have their roots in the social, political and economic beliefs of our society. When our values are in conflict out policies are in contention. The temptation then becomes overwhelming to seek to change social policies through attacks on the instruments through which the policies are implemented. Controversies over these policies too often are manifested in attacks on tests rather than upon the value choices in the policies that underlie the use of tests.

Bob continued by reading brief excerpts from the ETS guidelines. Again, however, since copies of the guidelines were distributed to all who attended the conference, they will not be summarized here.

In summing up the panel presentations, David Cohen drew attention to three issues. First, an issue raised by Barbara Lerner, is that in an area like testing, there is a question as to the extent it is possible to rely on strictly scientific standards. A second issue is the question of enforcement of any such standards. A third issue is why there seems to be such concern regarding standards--because there is more need for such standards, because such efforts appear fruitful in improving things or perhaps as a diversion from other issues like enforcement of already existing standards?

Following these remarks, the floor was opened up to comments and questions from the full group. David Prasse suggested that there likely is no answer to improving test use in particular kinds of statistics, and questioned what use such standards really are. Bill Pierce suggested that we might think about prospective use of standards--for example, we might get references to standards included in education legislation. Ann Cook observed, however, that people can spend great amounts of time on standards without even having much impact on practice. Several others observed that the panel session, quite

ironically seemed to have been mislabeled--that is, it ought not have been called truth-in-testing.

Also discussed was the issue of test security. Bob Solomon observed that when a test is used for remedial instruction or classroom planning, the test should be available to teachers, but when used for other purposes, for example selection, tests might have to be kept secure. Mildred Buck observed that it is important that parents and teachers have access to test content, adding that kids often miss out on alot of educaional content because there is too much focus in schools on tests. Lillian Weber observed that in her experience, test content is known by teachers and does affect curricula. Given that, she does not understand the rationale for test secrecy. If the content of a test is important educationally, and everyone agrees it is, then of course it should be open. Ann Cook observed that while tests do affect curriculum because they are so prevalent, it is often hard for teachers to help prepare kids for tests. For example, she cited one case in which teachers were unable to obtain a copy of a blank answer sheet so that they could be familiarized with the format of the test.

Bill Pierce observed that there is an important difference between ideal standards and minimum requirements and that this distinction needs to be made. Eric Gardner commenting on the issue of standards for test use, noted that a committee is currently working to develop standards for program evaluation, and the idea of treating the draft program evaluation standards at a future Consortium meeting was briefly discussed.

Panel on Minimum Competency Testing

The third conference panel, on minimum competency testing, was held on Tuesday morning, and was chaired by Walt Haney. The first panelist was Edith

Petrock of the Education Commission of the States. She briefly reviewed recent actions around the country concerning minimum competency testing. Her remarks were based mainly on a paper entitled "State Activity on Minimum Competency Testing," dated May 10, 1978. Since that paper was distributed at the conference, her remarks will not be summarized here.[10] Edith also discussed a policy paper of recommendation concerning minimum competency testing which is being prepared by ECS. A copy of that paper is also appended to this summary.

The second panelist was Peter Airasian of the American Educational Research Association. Peter discussed several issues regarding minimum competency testing which need to be considered before one even gets to the issue of measurement.[11] Specifically, he discussed four issues that need to be raised in the earliest stages of planning a minimum competency testing program: 1) the attempt to mandate or legislate pupil achievement; 2) instructional uncertainties; 3) the language of minimal competencies; and 4) the question of "Who is to be held accountable for learning?" After discussing each of these issues, Peter noted in conclusion that while they are related to the search for suitable standards and measurement techniques, these questions are important in their own right and should not be overlooked.

The third panelist on Tuesday morning was George Madaus who spoke on three measurement issues and consequences associated with minimum competency testing.[12]

---

[10] Copies of the paper are available from the Department of Research and Information, ECS, 1860 Lincoln Street, Suite 300, Denver, Colorado 80295.

[11] "Pre-measurement Issues in Minimal Competence Testing." A copy of this paper is appended to this summary. Hence, Peter's presentation will be only briefly summarized here.

[12] "Measurement Issues and Consequences Associated with Minimal Competency Testing." A copy of this paper is appended to this summary, so it is only briefly summarized here.

First, he noted that there are several historical precedents to the idea of minimum competency testing both in the United States and abroad. Second, he discussed some of the major consequences for students, teachers and the curriculum associated with testing programs designed to certify successful completion of a given level of education, consequences which he feels have direct relevance to minimum competency testing programs. Third, he outlined several technical issues of measurement that will inevitably arise when such programs become law.

The fourth participant on the minimum competency testing panel was Merle McClung of the Harvard Center on Law and Education.[13] Merle summarized three legal issues which seem to make minimum competency testing programs legally vulnerable. First, he noted that in many cases, such programs seem to be implemented without adequate phase-in periods. Second, he observed that often there seems to be a mismatch between what is tested and what is taught--raising the issue of whether minimum competency tests have instructional validity. Third, he noted the potential of minimum competency testing for racial and linguistic discrimination. To illustrate there problem areas, Merle drew on the example of the Florida minimum competency testing program for two reasons; first because Florida will probably be the first state in the nation to deny a high school diploma on the basis of a minimum competency test and second, because the state statute in Florida seems to be an example of the worst that can be found in minimum competency testing. Regarding phase-in, he noted that the Florida law was passed in 1976, so that would be the earliest time at which students might have been given notification of the new diploma requirement. Students were,

---

[13] No written version of Merle's presentation to the conference is available, but the gist of his comments are reported in three of his recently published articles: M. S. McClung, "Competency Testing: Potential for Discrimination," Clearinghouse Review, September 1977, pp. 439-48; M. S. McClung and D. Pullin, "Competency Testing and Handicapped Students," Clearinghouse Review, March 1978, pp. 922-27; M. S. McClung, "Are Competency Testing Programs Fair? Legal?" Phi Delta Kappan, February 1978, pp. 397-400.

of course, actually given notification after 1976, but whenever it was given, probably in 1977, it seems clear that the notification was not sufficiently early to allow for students to learn and for teachers to teach as they might have if they had received adequate notice of the new requirement. Regarding the match between testing and what is taught, Merle noted that minimum competency tests risk testing children on what they have never been taught. In order to overcome the arbitrariness of setting standards at particular levels, states seem to have enlisted broad public participation in the standard-setting process. But by including broad participation in the standard-setting process, Merle noted, states inevitably end up including some new or at least redefined educational goals. This results often in a mismatch between what is measured by the test and what is taught in school. This is especially true, Merle suggested, if the tests focus on functional literacy rather than just straight literacy which most school curricula are designed around. In considering legal ramifications, Merle had anticipated that it would be easy to find professional support for the idea that tests should not measure what has not been taught, but he discovered that it was very hard to find support for that idea in the professional literature. For example, he concluded after examining the technical testing standards' treatment of content validity, that this notion was inadequate. What is needed he suggested is a more specific type of content validity--what Merle calls instructional validity. The idea behind this term is simply that tests should not measure what has not been taught in schools. In the absence of professional standards on this point Merle noted that he and his colleagues are looking for a legal basis on which to raise the issue of instructional validity of minimum competency tests.

With respect to the potential of minimum competency tests for ethnic and linguistic discrimination, Merle noted that the basic legal principle is

the prohibition against carrying forward the effects of past racial discrimination. In lay terminology, this means that there is a prohibition against blaming the victim. Since there is very little law on minimum competency testing, the task for lawyers, Merle advised, is to bring the legal literature on racial discrimination to bear on this new area.

The fifth participant on the minimum competency panel was Hayes Mizell, of the Southeastern Public Education Program (SEPEP) of the American Friends Service Committee. Hayes noted initially that his group's background is in the civil rights experience and racial desegregation. He noted that SEPEP has become involved in minimum competency testing because of its potential impact in those areas. Hayes continued by presenting what he called his group's creed on minimum competency testing--that is a statement of where he and his colleagues are on this issue. First, they believe that the affects of minimum competency testing will be essentially negative because such programs devote resources to identifying and alledgedly remedying children's lack of competence, but do not devote comparable resources to identifying and building on the competencies and talents of children. Second, minimum competency testing programs perpetuate the notion that you can make children learn, and that if they do not respond to your efforts, it is appropriate to punish them. Third, they believe that minimum competency testing programs are based on ill-founded assumptions about the competence and willingness of most local school officials to provide effective remediation. Fourth, Hayes continued, we believe that there is an implicit misrepresentation in minimum competency programs that suggest that the attainment of certain minimum competencies will guarantee later success. Fifth, Hayes' group believes that minimum competency testing programs will be implemented not really with malice, but with good intentions,

which will nevertheless be harmful to the educational interests of minority children. Also, Hayes commented, his group believes that minimum competency testing will reinforce the notion that schools exist more for other people, for example for the business community and institutions of higher education, than for students or for parents. Hayes suggested that the SEPEP does believe that it is appropriate for local school systems to identify, after broad community consultation, the competencies the community feels are most important and to communicate those to parents and to students, early and regularly what those competencies are. Schools, moreover, should help students gain those competencies, as part of the overall school program, and periodically report on how students are doing in obtaining them. The SEPEP does not believe, however, that this should be done in a coercive way in a system based on fear. Their experience is that quality education occurs where teachers and administrators believe in the potential of all students, where educators develop and implement with energy and imagination effective programs to identify and build on that potential and where they initiate and implement with efforts to communicate to parents and children that they are valuable and respected participants in the educational process.

Unfortunately, however, minimum competency testing is, in Hayes' experience, largely a political response by legislators to parental and community unrest with respect to education. As such, it seems to reinforce sort of a "cool-whip" philosophy of educational reform--the idea that we can put a dollop of remedy on top and then basic problems will be resolved and everything will be more palatable.

Hayes reported that the SEPEP had prepared a Citizen's Introduction to Minimum Competency Programs for Students. Copies of the Introduction were

distributed at the conference.[14]

Hayes recounted also that over the year of its experience, his group has developed ideas about working with the federal establishment in order to promote the interests of its constituency. So the AFSC group has tried to foresee how existing federal programs relate to minimum competency testing programs. For example, Hayes was concerned with whether Title I funds might be used to pay for remedial efforts to help students who perhaps do not come up to required minimum comptency standards. So Hayes wrote to Commissioner Boyer to inquire about this possibility. Eventually, Commissioner Boyer wrote back to say essentially that in most cases Title I funds could not be used in such cases. Hayes also wrote to the Office of Civil Rights to inquire about a set of hypothetical issues in which minimum competency testing might have impact on individuals' civil rights. The OCR indicated that they were interested in such issues. For example, minimum competency testing may raise the issues of whether tests disproportionate impact on different minorities constitutes discrimination; how remediation relates to tracking and special education, what provisions there should be for students placed in special remediation classes to be retested and what the ramifications would be of schools awarding only an attendance certificate instead of a regular high school diploma. While the OCR evidenced interest in such issues, Hayes suggested that part of OCR's problem is that they know nothing about this issue and have apparently devoted very few staff members to minimum competency testing issues. In short, Hayes advised that there is a real need to get federal agencies to think about such issues ahead of time before they arise in specific cases in order to formulate sensible

---

[14] Additional copies of the Citizens Introduction are available for $1.50 per copy from the AFSC Southeastern Public Education Programs, 401 Columbia Building, Columbia, South Carolina, 29201.

policies about them.[15]

In conclusion, Hayes suggested that there is an enormous need for public education concerning minimum competency testing. Minimum competency testing need not be all bad, but there is a real need to get people to think about its limitations. Specifically, Hayes suggested, work needs to be done in the following areas:

-- What are the implications of existing laws of respective states for use of minimum competency tests with special groups, including ethnic and language minorities and handicapped children.

-- What are the outcomes of the implementation of minimum competency testing on those special groups, for example in terms of promoting drop-outs.

-- Someone should analyze the contents of minimum competency testing and publish a critical analysis of such tests, in a way that would contribute to public understanding of minimum competency testing.

Discussion following panel presentations was relatively brief due to the short time remaining before lunch. Vito Perrone commented that he is working with others to prepare a proposal to NIE to investigate the implementation of minimum competency testing programs in several states around the country, for example, maybe Oregon, Florida, Colorado, and Massachusetts. Several other people commented on the possibility of the National Consortium on Testing making some formal action regarding minimum competency testing or recommending certain policies regarding it. Following the brief discussion of the panel on minimum competency testing, the general session of the spring conference came to a close at about noon.

[Note: No action on the idea of the Consortium taking a position on minimum competency testing was accomplished at the business meeting on Tuesday afternoon since the group was unable to maintain a quorum for as long as would have been necessary to get to that issue.] 32

---

[15] Copies of Commissioner Boyer's letter regarding possible use of Title I funds in minimum competency testing programs, and a draft statement from OCR concerning civil rights implication of minimum comptency are avilable from Hayes Mizell at SEPEP, AFSC, 401 Columbia Building, Columbia, South Carolina 29201.

The Supreme Court and the APA, AERA, NCME Test Standards: Past References and Future Possibilities

Barbara Lerner

National Academy of Sciences

Washington, D.C.

DRAFT COPY - DO NOT QUOTE

Running head: The Supreme Court and the Test Standards

ABSTRACT

A survey of Supreme Court references to the Test Standards and to federal agency guidelines on testing over the past seven years suggests that the Court is beginning to rely on the Standards as an authoritative source for objective answers to technical, psychometric questions. Implications of this trend for the Joint Review Committee on Test Standards, its parent organizations, and the relationship between science and government are discussed. It is suggested that agency guidelines are inevitably based on a combination of research evidence and policy preferences, but that the Standards need not be, and will be most useful to the Court if they make scientific fidelity their sole touchstone.

The Supreme Court and the APA, AERA, NCME Test Standards: Past References and Future Possibilities

Barbara Lerner

National Academy of Sciences, Washington, D.C.

In this textually brief paper, I have chosen to focus on a single, legal-professional question which seems to me to be of overriding importance for the Joint Review Committee on Test Standards and its parent organizations, and not altogether without significance for the continuity of certain broad social and legal traditions which have nurtured the growth of both science and democracy in this country for some 200 years.

The question is this: when a legal challenge to test use arises, what set of standards will the Supreme Court of the United States rely upon in deciding whether that particular test use in that particular situation is acceptable in technical, psychometric terms?

Please note that the question posed here has to do with technical, psychometric standards only. The Court has not, will not, and, under our system of government, should not defer to any group of technical-scientific experts in deciding legal questions, in interpreting the intent of Congress as manifested--sometimes murkily--in legislation, or in choosing among the oftimes competing values enshrined in the majestic ambiguity of our Constitution.

Neither, however, under our system of government, should the Court ignore professional-technical experts, deciding essentially scientific questions by judicial fiat. Happily, although a few lower federal court

opinions contain pronouncements which move dangerously close to what might be described as the Lysenkoist position (Lerner, 1976), the Supreme Court of the United States seems to be cognizant of that danger and to be proceeding with appropriate caution.

Indeed, the Court is now proceeding so cautiously that no definitive answer to the question posed above can be given at this time, but a trend of considerable potential importance does seem to be developing: the Court seems to be moving away from reliance upon and deference to federal agency guidelines for this limited purpose and toward reliance upon and perhaps deference to professional standards as reflected in the APA, AERA, NCME Test Standards.

To allow you to judge for yourselves the existence and strength of this trend, I have attempted to excerpt and reproduce below every significant evaluative statement in every opinion of the Supreme Court by any of its nine Justices which refers to federal agency guidelines on test use and/or to the Standards. These excerpts are arranged in chronological order to highlight what seems to be a change over time in the aforementioned direction, and quoted at sufficient length to shed as much light as possible on the reasons the Court has thus far made explicit for its changing views on the different sets of guidelines and standards, their respective statuses and appropriate uses.

Assuming that this trend does, in fact, exist, whether it continues, accelerates or reverses itself is likely to depend, to a not inconsiderable extent, on the response of this Committee and of its parent organizations, now and over the course of the next few years. At this point in

time, what the Court needs--and seems to want--as circumstances force it to venture deeper and deeper into this unfamiliar and highly specialized area, is the psychometric equivalent of Gray's Anatomy: a dispassionate, lucid and practical set of technical-scientific standards, based upon but also limited by the best research evidence currently available and, as a consequence, capable of commanding clear majority support from recognized experts in the field of psychometrics, regardless of their views on what additional steps might or might not be desirable on the basis of social policy preferences.

Federal agency guidelines on testing inevitably incorporate and reflect such preferences. Their raison d'être is to obtain social action consistent with each agency's interpretation of its particular social mission (Comptroller General's Report to the Congress, 1978); scientific fidelity has not been and is not likely to be their sole touchstone. The most obvious manifestations of this basic fact are the conflicting sets of test guidelines issued by the several agencies and departments responsible for the interpretation and/or enforcement of the Civil Rights Act of 1964, especially since the 1972 amendments which made Title VII applicable to government as well as private employers (Novick, 1978).

Even if the long proposed Uniform Federal Guidelines are finally adopted by all of the agencies concerned, the result is likely to be a political compromise between competing social values, not a pure scientific product. As a result, the Court will still need--and may continue to want--an authoritative scientific source which is exactly that and only that.

I have chosen to highlight this issue at the outset because it seems to me inevitable that this Committee will choose whether it wishes to be responsive to what the Court needs and appears to want or whether it prefers to try to compete with the agencies by basing its recommendations on a combination of policy preferences and research evidence, rather than on evidence alone. That being the case, it seems preferable that our choice be made explicitly, at the start, and on the basis of a careful consideration of both the advantages and the disadvantages of either choice. We will, I think, structure our task quite differently, depending upon which alternative we choose.

To cite only the most obvious and general of examples, we will, if we choose to accept responsibility for the Gray's Anatomy role, do some, perhaps all, of the following things:

1) Place testing in its appropriate context as one of many possible decision-making devices, including information on the reliability and validity of its most common alternatives: interviews and supervisory ratings.

2) Focus as much on what is technically practical and possible as on what might be scientifically ideal and desirable, differentiating between those two poles as clearly as possible.

3) Take pains to point out both the strengths and the weaknesses of each of the three main approaches to validity, guarding against any tendency to downplay the weaknesses of one approach or to overplay the strenths of another.

4) Make corrections in those areas where research has moved ahead of the 1974 Standards and make deletions in those areas where the 1974 Standards went beyond and remain beyond the evidence accumulated to date.

5) Eschew the easy arrogance of hindsight, struggling to retain a sense of tradition and continuity, while still allowing room for necessary change, growth and development in a document which has survived a turbulent quarter century with considerable integrity albeit with, as yet, somewhat limited impact.

Excerpts from relevant Supreme Court cases, interspersed with a small amount of explanatory text follow. The intention is to highlight the fact that the Court is going through a period of struggle and transition in relation to the question this article focuses upon. Glib, stereotypic characterizations of this process as a struggle between "liberals" and "conservatives" miss the point: the struggle is within as well as between Justices. For those impatient with legal detail, a caution: friends of the Court, and the country, must care, not only about results in particular cases, but about the reasoning the Court uses to arrive at and/or to justify its conclusions.

Griggs v. Duke Power Co.

401 U.S. 424 (1971)

In this first testing case to reach the Supreme Court, Mr. Chief Justice Burger spoke for a unanimous bench of eight, joined in both his reasoning and his conclusion by all except Mr. Justice Brennan who took no part in the case. There is no mention of the Standards, only

of agency guidelines on testing, which are said to be "entitled to great deference" as follows:

> The Equal Employment Opportunity Commission, having enforcement responsibility, has issued guidelines interpreting §703(h) [of the Civil Rights Act of 1964] to permit only the use of job-related tests.[9] The administrative interpretation of the Act by the enforcing agency is entitled to great deference. See, e.g., United States v. City of Chicago, 400 U.S. 8 (1970); Udall v. Tallman, 380 U.S. 1 (1965); Power Reactor Co. v. Electricians, 367 U.S. 396 (1961). Since the Act and its legislative history support the Commission's construction, this affords good reason to treat the guidelines as expressing the will of Congress.
>
> (pp. 433-434).

Albemarle Paper Co. v. Moody

422 U.S. 405 (1975)

Four years after Griggs, Albemarle, a case with very similar facts reached the Court. The immediate result--the judgment--was also essentially the same, but the process by which the Court reached it and, hence, the predictive implications of the case, were very different, for reasons I have attempted to spell out elsewhere (Lerner, 1976).

This time, Mr. Justice Stewart delivered the opinion of the Court, unqualifiedly joined in both his reasoning and his result by only three of his colleagues: Justices Douglas, Brennan and White. The same agency guidelines are still said to be "entitled to great deference," but now an attempt is made to bolster that claim by taking judicial notice of the

Standards and stressing the relationship of the former to the latter, and the presumed congruence between them.

Justices Marshall and Rehnquist found it necessary to file separate, concurring opinions. Mr. Justice Blackmun filed an opinion concurring in the judgment only. Mr. Chief Justice Burger also filed a separate opinion, concurring in part and dissenting in part, insisting that the Court's reliance on his own earlier "great deference" statement was "misplaced," objecting to the Court's "wooden application" of the agency guidelines, and highlighting the discrepancies between those guidelines and the Standards. Mr. Justice Powell took no part in the case. Relevant excerpts follow.

Stewart, J., for the Court

> The EEOC has issued 'Guidelines' for employers seeking to determine, through professional validation studies, whether their employment tests are job related. 29 CFR pt. 1607. These Guidelines draw upon and make reference to professional standards of test validation established by the American Psychological Association.[29] The EEOC Guidelines are not administrative 'regulations' promulgated pursuant to formal procedures established by the Congress. But, as this Court has heretofore noted, they do constitute '[t]he administrative interpretation of the Act by the enforcing agency,' and consequently they are 'entitled to great deference.' Griggs v. Duke Power Co., 401 U.S., at 433-434. See also Espinoza v. Farah Mfg. Co., 414 U.S. 86, 94 (1973). ...

(pp. 430-433)

In Griggs v. Duke Power Co., supra, the Court left open 'the question whether testing requirements that take into account capability for the next succeeding position or related future promotion might be utilized upon a showing that such long-range requirements fulfill a genuine business need.' 401 U.S., at 432. The Guidelines take a sensible approach to this issue, and we now endorse it.

(pp. 433-434)

Blackmun, J., Concurring in the Judgment

I cannot join, however, in the Court's apparent view that absolute compliance with the EEOC Guidelines is a sine qua non of pre-employment test validation. The Guidelines, of course, deserve that deference normally due agency statements based on agency experience and expertise. Nevertheless, the Guidelines in question have never been subjected to the test of adversary comment. Nor are the theories on which the Guidelines are based beyond dispute. The simple truth is that pre-employment tests, like most attempts to predict the future, will never be completely accurate. We should bear in mind that pre-employment testing, so long as it is fairly related to the job skills or work characteristics desired, possesses the potential of being an effective weapon in protecting equal employment opportunity because it has a unique capacity to measure all applicants objectively on a standardized basis. I fear that a too-rigid application of the EEOC Guidelines will leave

the employer little choice, save an impossibly expensive and complex validation study, but to engage in a subjective quota system of employment selection. This, of course, is far from the intent of Title VII.

(p. 449)

Burger, C.J., Concurring in Part and Dissenting in Part

The Court's treatment of the testing issue is equally troubling. Its entire analysis is based upon a wooden application of EEOC Guidelines which, it says, are entitled to 'great deference' as an administrative interpretation of Title VII under Griggs v. Duke Power Co., 401 U.S. 424 (1971). The Court's reliance upon Griggs is misplaced. There we were dealing with Guidelines which state that a test must be demonstrated to be job related before it can qualify for the exemption contained in §703 (h) of Title VII, ... as a device not 'designed, intended or used to discriminate....' Because this interpretation of specific statutory language was supported by both the Act and its legislative history, we observed that there was 'good reason to treat the guidelines as expressing the will of Congress....'

In contrast, the Guidelines upon which the Court now relies relate to methods for proving job relatedness; they interpret no section of Title VII and are nowhere referred to in its legislative history. Moreover, they are not federal regulations which have been submitted to public comment and scrutiny as

required by the Administrative Procedure Act.[3] Thus, slavish adherence to the EEOC Guidelines regarding test validation should not be required. Those provisions are, as their title suggests, guides entitled to the same weight as other well-founded testimony by experts in the field of employment testing.

(pp. 451-452)

(Footnote 3:)

Such comment would not be a mere formality in light of the fact that many of the EEOC Guidelines are not universally accepted. For example, the Guideline relating to 'differential validation' upon which the Court relies in this case ... has been questioned by the American Psychological Association. See United States v. Georgia Power Co., 474 F.2d 906, 914 n.8 (CA5 1973).

(p. 452)

Washington v. Davis

426 U.S. 229 (1976)

Facts, reasoning and results are all different in this case (Lerner, 1976;1978). So, too, is the attitude of the majority towards the relationship between the Standards and agency guidelines, with the former now seeming to take something approaching preeminence over the latter, although disparities between them are still being downplayed. Mr. Justice White delivered the opinion of the Court, joined by Chief Justice Burger and by Justices Blackmun, Powell and Rehnquist. Mr. Justice Stewart joined in parts I and II of the opinion, creating, in effect, a

5 2/3 man majority for the Court's reasoning. Mr. Justice Stevens filed a separate, concurring opinion. Mr. Justice Brennan filed a dissent, joined by Mr. Justice Marshall. The two dissenters are now alone in ignoring the Standards, insisting that agency guidelines should govern absolutely, even though this particular case was--for temporary technical reasons--not within the agency's jurisdiction.

White, J., for the Court

(Footnote 13:)

It appears beyond doubt by now that there is no single method for appropriately validating employment tests for their relationship to job performance. Professional standards developed by the American Psychological Association in its Standards for Educational and Psychological Tests and Manuals (1966), accept three basic methods of validation: "empirical" or "criterion" validity (demonstrated by identifying criteria that indicate successful job performance and then correlating test scores and the criteria so identified); "construct" validity (demonstrated by examinations structured to measure the degree to which job applicants have identifiable characteristics that have been determined to be important in successful job performance); and "content" validity (demonstrated by tests whose content closely approximates tasks to be performed on the job by the applicant). These standards have been relied upon by the Equal Employment Opportunity Commission in fashioning its Guidelines on Employee Selection Procedures, 29 CFR pt. 1607 (1975), and have been

judicially noted in cases where validation of employment tests has been in issue. See, e.g., Albemarle Paper Co. v. Moody, 422 U.S. 405, 431 (1975); Douglas v. Hampton, 168 U.S. App. D.C., at 70,512 F.2d, at 984; Vulcan Society v. Civil Service Comm'n, 490 F. 2d 387, 394 (CA2 1973).

(p. 247)

Brennan, J., Dissenting

Today's decision is also at odds with EEOC regulations issued pursuant to explicit authorization in Title VII. 42. U. S. C. §2000e-12(a). Although the dispute in this case is not within the EEOC's jurisdiction, as I noted above, the proper construction of Title VII nevertheless is relevant. Moreover, the 1972 extension of Title VII to public employees gave the same substantive protection to those employees as had previously been accorded in the private sector, Morton v. Mancari, 417 U.S. 535, 546-547 (1974), and it is therefore improper to maintain different standards in the public and private sectors. Chandler v. Roudebush, 425 U.S. 840, 864 (1976). See n.2, supra.

As with an agency's regulations, the construction of a statute by the agency charged with its administration is entitled to great deference. Trafficante v. Metropolitan Life Ins. Co., 409 U.S. 205, 210 (1972); Udall v. Tallman, 380 U.S., at 16; Power Reactor Co. v. Electricians, 367 U.S. 396, 408 (1961).

The deference due the pertinent EEOC regulations is enhanced by the fact that they were neither altered nor disapproved when Congress extensively amended Title VII in 1972.[8] Chemehuevi Tribe of Indians v. FPC, 420 U.S. 395, 410 (1975); Cammarano v. United States, 358 U.S. 498, 510 (1959); Allen v. Grand Central Aircraft Co., 347 U.S. 535, 547 (1954); Massachusetts Mut. Life Ins. Co. v. United States, 288 U.S. 269, 273 (1933). These principles were followed in Albemarle--where the Court explicitly endorsed various regulations no fewer than eight times in its opinion, 422 U.S., at 431-436[9]--and Griggs, 401 U.S., at 433-434.

(pp. 263-264)

(Footnote 8:)

Still another factor mandates deference to the EEOC regulations. The House and Senate committees considering the 1972 amendments to Title VII recognized that discrimination in employment, including the use of testing devices, is a "complex and pervasive phenomenon." S. Rep. No. 92-415, p. 5 (1971); H. R. Rep. No. 92-238, p. 8 (1971). As a result, both committees noted the need to obtain "expert assistance" in this area. S. Rep. No. 92-415, supra, at 5; H. R. Rep. No. 92-238, supra, at 8.

General Electric Co. v. Gilbert

429 U.S. 125 (1976)

This is not a testing case; it is included here because it contains many clearly evaluative statements about EEOC Guidelines. Mr. Justice Rehnquist delivered the opinion of the Court, joined by Justices Burger,

White and Powell. Mr. Justice Stewart filed a concurring statement. Mr. Justice Blackmun filed a statement concurring in part. Mr. Justice Brennan filed a dissenting opinion, joined by Mr. Justice Marshall. Mr. Justice Stevens filed a separate dissenting opinion.

Rehnquist, J. for the Court

> We are told, however, that this analysis of the congressional purpose underlying Title VII is inconsistent with the guidelines of the EEOC, which, it is asserted, are entitled to "great deference" in the construction of the Act, Griggs, 401 U.S., at 433-434; Phillips v. Martin Marietta Corp., 400 U.S. 542, 545 (1971) (Marshall, J., concurring). The guidelines upon which respondents rely most heavily was promulgated in 1972, and states in pertinent part: ...
>
> (p. 140)
>
> In evaluating this contention it should first be noted that Congress, in enacting Title VII, did not confer upon the EEOC authority to promulgate rules or regulations pursuant to that Title. Albemarle Paper Co. v. Moody, 422 U.S. 405, 431 (1975).[20] This does not mean that EEOC guidelines are not entitled to consideration in determining legislative intent, see Albemarle, supra; Griggs v. Duke Power Co., supra, at 433-434; Espinoza v. Farah Mfg. Co., 414 U.S. 86, 94 (1973). But it does mean that courts properly may accord less weight to such guidelines than to administrative regulations which Congress has declared shall have the force of law, see Standard Oil Co. v. Johnson, 316 U.S.

481, 484 (1942), or to regulations which under the enabling statute may themselves supply the basis for imposition of liability, see, e.g., §23 (a), Securities Exchange Act of 1934, 15 U.S.C. §78w (a). The most comprehensive statement of the role of interpretative rulings such as the EEOC guidelines is found in Skidmore v. Swift & Co., 323 U.S. 134, 140 (1944), where the Court said:

"We consider that the rulings, interpretations and opinions of the Administrator under this Act, while not controlling upon the courts by reason of their authority, do constitute a body of experience and informed judgment to which courts and litigants may properly resort for guidance. The weight of such a judgment in a particular case will depend upon the thoroughness evident in its consideration, the validity of its reasoning, its consistency with earlier and later pronouncements, and all those factors which give it power to persuade, if lacking power to control."

The EEOC guideline in question does not fare well under these standards. It is not a contemporaneous interpretation of Title VII, since it was first promulgated eight years after the enactment of that Title. More importantly, the 1972 guideline flatly contradicts the position which the agency had enunciated at an earlier date, closer to the enactment of the governing statute. ...

(pp. 141-142)

We are not reduced to such total abdication in construing the statute. The EEOC guideline of 1972, conflicting as it does with earlier pronouncements of that agency, and containing no suggestion that some new source of legislative history had been discovered in the intervening eight years, stands virtually alone. Contrary to it are the consistent interpretation of the Wage and Hour Administrator, and the quoted language of Senator Humphrey, the floor manager of Title VII in the Senate. They support what seems to us to be the "plain meaning" of the language used by Congress when it enacted §703 (a) (1).

(p. 145)

Brennan, J., Dissenting

As a matter of law and policy, this is a paradigm example of the type of complex economic and social inquiry that Congress wisely left to resolution by the EEOC pursuant to its Title VII mandate. See H. R. Rep. No. 92-238, p. 8 (1972). And, accordingly, prior Title VII decisions have consistently acknowledged the unique persuasiveness of EEOC interpretations in this area....

(pp. 155-156)

For me, the 1972 guideline represents a particularly conscientious and reasonable product of EEOC deliberations and, therefore, merits our "great deference." Certainly, I can find no basis for concluding that the guideline is out of step with congressional intent.

(pp. 157-158)

NEA v. State of South Carolina

Nos. 77-422 & 77-543. Decided January 16, 1978

In this case, the Supreme Court affirmed the judgment of the 3-judge district court panel below, speaking only four words: "The judgment is affirmed." Mr. Justice White dissented, joined by Mr. Justice Brennan. Mr. Justice Marshall and Mr. Justice Blackmun took no part.

When the Court affirms without issuing an opinion, it is impossible to be certain of the exact basis for its affirmance or the extent of its agreement with the reasoning of the Court below. Nonetheless, the following language from the district court's opinion seems especially relevant in the context of this memorandum. See especially footnote 20 infra, and note that the phrase 'relegated to a footnote' is foreign to lawyers who, unlike psychologists, tend to use footnotes heavily, and to regard them at least as seriously as they do the text--and sometimes more so.

United States v. State of South Carolina

DC S.C. Civil Action No. 75-1610

April 14, 1977

The design of the validity study is adequate for Title VII purposes. The Supreme Court made clear once again in Washington v. Davis that a content validity study that satisfies professional standards also satisfies Title VII. 426 U.S. at 247, n. 13. The defendants called as an expert witness Dr. Robert M. Guion,

the principal author of Standards for Educational and Psychological Tests published by the American Psychological Association and a nationally recognized authority in the field of testing and measurement who testified in an unqualified fashion that in his expert opinion the ETS study design met all of the requirements of the APA Standards, the Division 14 Principles,[19] and the EEOC Guidelines.[20]

[19] Division of Industrial Organizational Psychology (Division 14), American Psychological Association, Principles for the Validation and Use of Personnel Selection Procedures (1975).

[20] Guidelines for Employee Selection Procedures, 29 C.F.R. 1607 (1970). To the extent that the EEOC Guidelines conflict with well-grounded expert opinion and accepted professional standards, they need not be controlling. General Electric Co. v. Gilbert, ____ U.S. ____, 97 S.Ct. 401 (1976).

Reference Note

1. Problems with federal equal employment opportunity guidelines on employee selection procedures need to be resolved, Report to the Congress by the Comptroller General of the United States, February 2, 1978 (FPCD-77-54).

REFERENCES

Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975).

General Electric Co. v. Gilbert, 429 U.S. 125 (1976).

Griggs v. Duke Power Co., 401 U.S. 424 (1971).

Lerner, B. Washington v. Davis: Quantity, quality and equality in employment testing. In P. Kurland (Ed.), The Supreme Court Review (1976 vol.). Chicago: University of Chicago Press, 1977.

Lerner, B. Equal protection and external screening: Davis, De Funis, and Bakke. (Proceedings of 1977 Invitational Conference.) Princeton, N.J.: Educational Testing Service, 1978.

National Education Association v. State of South Carolina, Nos. 77-422 and 77-543. Decided January 16, 1978.

United States v. State of South Carolina, DC S.C. Civil Action No. 75-1610, April 14, 1977.

Novick, M.R. The Influence of the Law on Professional Measurement Standards. (Proceedings of the 1977 Invitational Conference.) Princeton, N.J.: Educational Testing Service, 1978.

Standards for educational and psychological tests. Washington, D.C.: American Psychological Association, 1974.

Washington v. Davis, 426 U.S. 229 (1976).

FOOTNOTES

This article is based upon a memorandum requested by Melvin R. Novick, Chairman of the Joint Review Committee on Test Standards and submitted to my colleagues on that Committee at our March, 1978 meeting in Washington, D.C. Their thoughtful and generous responses helped me to clarify the text and encouraged me to offer it for publication at this time, but responsibility for the final product is strictly my own. I would like to thank all of my fellow Committee members--Robert Dugan, Ronald R. Edmonds, Ronald K. Hambleton, George Madaus, Jay Millman, Barbara Pedulla and Daniel Stufflebeam--and, especially, Melvin R. Novick, Chairman.

This material was also presented as part of a symposium at the March, 1978 American Educational Research Association Convention in Toronto, Canada. Appreciation is due to all participants in that symposium for their helpful comments.

In addition, I am indebted to Alexandra Wigdor, research associate for the Committee on Ability Testing for aid in compiling and double-checking the accuracy of all quoted excerpts from Supreme Court opinions.

Requests for reprints should be sent to Barbara Lerner, National Academy of Sciences, 2101 Constitution Avenue, Washington, D.C. 20418.

April 19, 1978

POLICY ISSUES IN MINIMUM COMPETENCY STANDARDS AND TESTING

I. OVERVIEW OF MINIMUM COMPETENCY ISSUES

A. Introduction

A major educational concern in the 1970's is a belief by many people that students are not acquiring the basic skills necessary to be successful in today's society. For those who share this concern, a common focus has been the proposal that minimum standards of skills and knowledge be identified as the basis for high school graduation. The reformers are suggesting that the present educational system which requires 12 years of attendance and completion of a predetermined set of credits does not assure the acquisition of the needed competencies. Their proposal for minimum standards generally involves selecting and defining competencies they consider necessary for success, establishing minimum levels of proficiency, and developing tests to determine whether or not the standards are being met. The hope is that a program of minimum competency testing will result in clearer definitions of educational goals and give added meaning to the high school diploma.

A majority of the states now have some form of competency testing as a result of action by the state legislature or state board of education. Among the 30 states which had taken action by November 15, 1977, policy was directed toward such diverse purposes as setting standards for grade to grade promotion and/or high school graduation, collecting information for evaluating the performance of schools or specific programs, or allocating funds. Some states specify that many decisions are to be made at

the local level while others centralize similar decisions at the state level. Some of the legislation is directed at only the high school student while some involves students at all levels. Generally, minimum competency testing of individuals is separate from state assessment programs, which seek to determine whether the learning of students, on the average, is improving or not.

Through the monitoring of state activity and study of related issues, the Education Commission of the States (ECS) has identified minimum competency standards and testing as an important but complex issue in the development of education within the United States. The high level of requests for information from constituents encouraged ECS to publish a series of Updates (January 1976-November 1977) to keep people informed about the type of action that is rapidly emerging. The co-sponsorship of conferences in September and October, 1977, allowed political and educational leaders to share perspectives and identify concerns and issues to be addressed. The ECS Policy Committee selected minimum competency as an issue for concentrated attention in 1977-78. Through these activities, the issue of minimum competency testing has been explored, and information to shape policy recommendations has been developed.

B. Political Considerations

The single label "minimum competency" presently serves to unify the efforts of groups and individuals who support programs which vary significantly in both intent and design. The laws and regulations that have been enacted provide evidence that some programs are designed to evaluate the performance of schools and others the performance of students. Some plans are designed to provide accountability and to save money, while others are directed primarily toward the improvement of instruction and remediation. While it is easy to misjudge a movement when it is first emerging, some of the diverse interests and their reasons for supporting minimum competency as an educational reform are definable:

1. A number of groups and individuals are searching for something which they feel has been lost. Convinced that the quality of education is on the decline, they point to several signs that students are not sufficiently equipped to survive in the complex post-industrial society they are entering. Some of the signs they cite as indicators that students are ill-prepared to adequately read, write, compute, and complete other important but more ambiguously defined tasks necessary for success are:

Slow steady deleted →

- Decline in both the verbal and mathematical sections of the Scholastic Aptitude Test (SAT)
- High enrollment in remedial courses for students entering college
- Lawsuits by high school graduates charging that diplomas were awarded without providing the skills to perform adequately
- Complaints from parents and employers that students have not been challenged to develop work habits or competencies required in the world of work
- Requests from some students that schools be more rigorous in demands for excellence

Generally, those who express these concerns feel that the deficiencies can be corrected by returning to the high standards of performance and discipline which many believe were characteristic of a past era, when a smaller percentage completed high school and a much smaller percentage went on to college. "Back-to-the-basics" has become a unifying theme for those who hope schools can be more like they believe they were in the past. Some are suggesting the implementation of a European-style curriculum and credentialing process to restore high standards.

2. Paralleling the concern about the decline in the quality of education are equally serious concerns about the sharply rising costs of

schooling. A desire to limit the increases, if not cut the levels of spending, has been widely discussed by state legislatures, which, on a national average, provide over 40% of the funds for education, and by local taxpayers who are also painfully aware of the sharp increases in tax assessments. To those responsible for allocating the funds, it is incomprehensible that test scores have declined during the same period of time (1966-1976) that per-pupil expenditures (corrected for inflation) have increased 47% on a national average.[1] Pointing to the limited success of Title I and compensatory education programs as well as the discrepancies between test scores and high levels of funding, some legislators and taxpayers are unwilling to concede that more money is the answer. Rather, they want an accounting for the money presently spent. The feeling is that both students and schools must make a more serious effort to achieve at higher levels consistent with the funds already allocated.

Many accountability advocates are hopeful that the push for standards will bring a more efficient system of educating students, unlike many of the ineffective accountability efforts of the past decade. Within this group, the establishment of minimum standards is viewed as a possible mechanism for prohibiting further increases in educational costs. Included in the efficient operation might be early exit for students who can demonstrate proficiency at minimum levels and choose to leave school, excusing students from some courses if they can demonstrate proficiency, and replacement of teachers and other personnel who do not demonstrate productiveness. For other accountability advocates, there is hope that money can be saved

---

[1]Figures for expenditures per pupil in average daily attendance (cited in National Education Association, Estimates of School Statistics, 1976, p. 18) were divided by the change in the consumer price index from 1966 to 1976.

through narrowing the focus of the school by giving high priority to those programs found to be integral to basic skill development and eliminating programs which do not contribute to specified goals. A corollary to making schools more efficient is the belief that students should work harder and take advantage of opportunity provided at public expense.

3. Others are interested in minimum competency testing because they hope to define education differently than in the past. Many disagree with some of the premises outlined above, and suggest that the reports of test score decline have been exaggerated. They point to the findings of mixed achievement in many academic areas by the National Assessment of Educational Progress (NAEP) and to the consistent scores on the Preliminary SAT among high school juniors as evidence that students are doing as well as in the past. Still, they feel that students need better and different training and preparation for the future than the schools have traditionally provided.

Among the leaders of this school of thought are advocates for competency-based education who are interested in some far-reaching reforms in educational structures. Their goal is to define instruction more in terms of specific outcomes related to life roles (worker, consumer, parent, friend, etc.) rather than through traditional time-based curriculum. Such a reform is suggested as the basis for providing students with the environment and opportunity to acquire a set of skills appropriate for their needs and interests.

4. Advocates for special school populations are also interested in minimum competency testing. Minority groups are hopeful that the establishment of standards will force school systems to find methods to provide basic

skills to the large numbers of students from culturally different backgrounds who are not succeeding in the present system. While strong interest in improved skill development has been expressed, minorities also express concern about the unfair use of tests and standards against students as evidenced by tracking or denial of opportunity.

The frustrations and concerns of the various interests described above are far from uniform and have created varying political pressure on both policy makers and practitioners at both the state and local level. Those sensitive to the demands for setting standards find it difficult to respond to the diverse and sometimes competing interest groups who are demanding change for very different purposes. While it is not easy or popular to resist the pressure to do something, it is very difficult to design a plan which will satisfy those whose primary motive is to save money with the same plan as those who desire high levels of proficiency for all students at any cost. The setting of minimum standards, which at first glance appears to be politically appealing, becomes very complex with the convergence of various ideas and interests, as decisions are sought regarding the design and implementation.

C. Technical Considerations

While political considerations have captured much of the attention among educational decision makers, many practitioners are suggesting that technical issues also need attention. Some people are critical of those legislative plans which they feel have overlooked some serious technical considerations in the haste to do something to appease unhappy constituents. Experts in evaluation point out that tests are not perfect. Too often, a student's score on a test is interpreted and summarized without sufficient regard for what the test is measuring and how well it is succeeding.

Because there are many problems inherent in test construction, policy makers have an obligation to draw upon the best technical knowledge that is available as part of the process of developing state or local approaches to minimum competency testing.

The problems of building a test to determine minimum competency includes issues related to format, test conditions, validity (How well does the test measure what it was designed to measure?), and reliability (How consistently does the test measure that?). The most easily constructed, economical, and available tests involve pencil and paper, often in a multiple choice format. While these tests are successful in measuring basic cognitive skills, there are many objectives of minimum competency programs which involve the performance of complex skills in areas which are difficult to quantify. Yet, tests of application and simulation to evaluate the more complex skills are expensive and time-consuming to develop, administer, and interpret. Understanding the limitations and proper uses for tests should help policymakers insure that tests are not overused or underused as a source of information.

Another technical question relates to the search for methods of determining a cut-off point that is not arbitrary and unfair. Should the cut-off point be the same for all pupils? What are the skills that all pupils must master? While research continues, experts point out that no formula exists for objectively selecting the "notch" indicating mastery or need for remediation. Thus, standards for proficiency must be judgementally determined and should be subject to public scrutiny and challenge.

Cultural bias is another technical concern. Because the experiences of students vary, it is difficult to design a test without including some

questions which provide an advantage to some students because the content relates more directly to their life experience. The development of questions which relate to the common core of experiences of all children taking the test should be encouraged.

II. POLICY ISSUES

A. Discussion of Issues

Responsible policy at any level will involve comprehensive and cooperative study by educational and political leaders of the political and technical realities as part of the search for commonly acceptable solutions for improving basic skills. The primary thrust must be for policy which supports improvements in the teaching/learning process. A test by itself is not a solution to learning problems. Effective programs must include attention to the instruction that precedes and follows the test. The goal of improving student achievement should be a primary focus of discussions of testing and standard setting.

To encourage a positive and permanent impact on the instruction of basic skills, policymakers should avoid the tendency of other accountability-laden proposals to respond to public criticism and demands for change with a mechanical solution that has little or no relationship to how students learn.

The decision to, or not to, implement a minimum competency testing program should involve a weighing of the positive and negative consequences of either decision. The following list[2] of possible outcomes can serve as a

---

[2] This list is from Competency Tests and Graduation Requirements, National Association of Secondary School Principals, Reston, VA. 20091, 1976.

starting point for analyzing whether or not minimum competency testing would be a good approach to improve the attainment of basic skills.

Possible Positive Outcomes:

- The question, "What is a high school education?" must be squarely faced.
- The statements required for each course will likely result in carefully organized teaching and carefully designed sequential learning.
- Slow learners and underachievers will likely receive direct and immediate attention.
- Courses of study will likely be revised to correct identified deficiencies.
- Subjects leading to the development of competencies will receive additional emphasis.
- Alternatives and options not requiring attendance in class will likely be broadened.
- The senior year may gain more holding power because of a new focus upon requirements and options.
- The community will know the minimum performance required in specific subject areas for the diploma.

Possible Negative Outcomes:

- Confusion over the meaning of a high school diploma will continue if each district identifies its own level of competencies and performance indicators.
- The emphasis on pragmatic and practical competencies may result in erosion of liberal education.

- The emphasis on measurable outcomes could result in less attention to outcomes which are difficult to measure.
- The record-keeping system could become burdensome to teachers and administrators.
- The conflict between "humaneness" and "accountability" may be intensified as criteria are established and clarified.
- Community disagreement may arise over the nature and difficulty of competencies.
- Dropouts could increase depending upon the level of the minimum competencies.

The complexity of issues and range of unanswered questions must be directly and deliberately addressed. Policy should encourage research and the incorporation of findings into any decision to move toward minimum competency testing. If the conclusions of research papers completed by scholar/practitioners for the series of conferences sponsored by ECS and NIE (National Institute of Education) are any indication, there are many concerns which should cause people to move carefully toward program design and implementation. States should examine the differences that existing programs are making when embarking on a plan to improve achievement.

B. ECS Policy Positions

ECS has adopted the following policies in regard to minimum competency standards and testing for use by education policymakers at federal, state and local levels. As new knowledge is available from research and experience, these policies may be amended.

1. ECS SUPPORTS STATE EFFORTS TO IMPROVE EDUCATIONAL ACHIEVEMENT, AND TO INCREASE THE EDUCATIONAL ATTAINMENT OF UNDERACHIEVING STUDENTS TO LEVELS

WHICH WILL ENABLE THEM TO PARTICIPATE EFFECTIVELY IN SOCIETY, INCLUDING OBTAINING AND RETAINING A JOB. SETTING STANDARDS FOR GRADUATION, GRADE TO GRADE PROMOTION OR PLACEMENT IN PROGRAMS AND SPECIFYING METHODS FOR DETERMINING WHEN THOSE STANDARDS HAVE BEEN MET ARE COMMON PRACTICES IN EDUCATION. SOME STANDARDS ARE SET BY THE STATE, SOME BY THE LOCAL SCHOOL DISTRICT OR SCHOOL. TEACHER-MADE TESTS, DEPARTMENT-WIDE EXAMINATIONS, STANDARDIZED TESTS, MINIMUM COMPETENCY TESTS, TEACHER JUDGEMENT, DEMONSTRATION OF MASTERY BY STUDENT PERFORMANCE ARE SOME OF THE METHODS USED TO DETERMINE WHEN PUPILS HAVE MET THE STANDARDS. USE OF MINIMUM STANDARDS AND MINIMUM COMPETENCY TESTS TO DETERMINE WHEN THE STANDARDS HAVE BEEN MET IS ONE OPTION. THERE IS NOT, HOWEVER, A BODY OF KNOWLEDGE OR RESEARCH DATA TO SUPPORT A CONCLUSION THAT THIS METHOD IS BETTER THAN OTHERS. THEREFORE, ECS ADVOCATES THAT STATES AND LOCAL DISTRICTS ADOPT A COMPREHENSIVE PLAN FOR SETTING STANDARDS AND ESTABLISHING CRITERIA FOR DETERMINING WHEN THOSE STANDARDS HAVE BEEN MET. SUCH A PLAN SHOULD AVOID THE SELECTION OF STANDARDS THAT ARE TOO NARROW AND THE RELIANCE ON A SINGLE ASSESSMENT TECHNIQUE.

Important Considerations:

- Minimum competency testing and standards are one part of a larger national concern about achievement in education. Among those dedicated to and actively working for improvement in learning and achievement in the basic skills, there are strongly held, often widely differing views as to the promise that minimum competency testing holds for attainment of these objectives.
- A variety of methods are presently available for assessing students, determining standards, and establishing quality controls within curricular programs. Leaders at both the local and state level are exploring other options in addition to competency testing for

improving student achievement. The diverse approaches include use of state-wide assessment data to improve program, management by objectives to evaluate instruction, revision of high school graduation requirements, individualized learning programs, and targeting of individual learning problems for specialized help. The decision not to set minimum standards through testing is often made because of the complex technical and political issues which must be addressed when designing and implementing a minimum competency testing program.

2. IN ANY CONSIDERATION OF THE ESTABLISHMENT OF MINIMUM COMPETENCY TESTING AND STANDARDS, THE PURPOSE FOR THE TESTING AND STANDARDS SHOULD BE CLEARLY DETERMINED.

Important Considerations:

- The primary reason that minimum competency testing and standards are being implemented is to ensure a minimum level of achievement by pupils. In designing minimum competency legislation or policies, special attention should be given to see that this purpose will be achieved. For the test results to be helpful, it will be necessary to understand the limitation and uses for tests.
- Minimum competency standards must not be confused with the goals and objectives of education for the public schools. Goals and objectives should be designed to foster maximum development of each pupil. Minimum standards are only one aspect of the total education program. The purpose of minimum competency standards is to set a lower limit of acceptable achievement in designated basic skills for pupils. Minimum standards are not adequate

measures of the output or effectiveness of individual teachers, the adequacy of a total school program or the full range of an individual's skills and abilities.

- One promising use of minimum competency testing for ensuring achievement in the basic skills is to use these tests as an early warning device or for diagnostic purposes early in the elementary grades. Low achieving pupils should then be placed in instructional programs designed to improve their achievement in those skill areas in need of improvement. Children should be taught those skills that the tests indicate they do not know, not recycled through a program that does not effectively meet the individual needs.
- The use of a single test of competency to determine student qualifications for graduation or grade to grade promotion is an inadequate measure. Given the limitations of a single test and the importance of such decisions to the affected individual, other evidence such as the teachers' observations and judgement should play a part in these decisions.
- Since the source of many of the problems related to low achievement have family and societal origins, the school will not be able to solve the problems without the help of the family and community.[3] A partnership which defines the role of school, home, and community should be the basis of a minimum competency program.

---

[3] A blue ribbon panel of experts headed by Willard Wirtz and commissioned by the College Entrance Examination Board (CEEB) found lower standards to be only one of the many explanations for the decline in test scores. Other causes included societal changes, large numbers of hours spent watching television, and increasing numbers of students living in unstable family situations or working at full or part time jobs while attending school

3. DIAGNOSIS OF LEARNING PROBLEMS OF PUPILS WHO DO NOT ATTAIN MINIMUM STANDARDS AND IMPLEMENTATION OF INSTRUCTION PROGRAMS THAT FOCUS ON THEIR PARTICULAR DEFICIENCIES SHOULD BE COMPONENTS OF ANY MINIMUM COMPETENCY PROGRAM.

Important Considerations:

- The enormous range of differences in individual abilities must be recognized. There will always be some who fall below the minimum standards in spite of the best effort of the school and student. Minimum competency laws or policies should recognize individual differences and make provisions which are educationally productive for those students who cannot pass the test.
- Frequently, minimum competency policies call for retention or remediation for pupils who score below the minimum. The evidence from research is that routine repetition of a grade is not an effective means for improving achievement. Remediation in the basic skills is sometimes effective and sometimes not, often because the rates and modes in which children learn may not be clearly identified. Remediation is more likely to be effective if started in the early grades and if the instruction is targeted at the particular learning level of the individual. Therefore, staff expertise and funds should be provided which will make it possible for schools to design and implement remedial programs that are based on individual needs and learning styles.

4. THE ROLES OF ALL LEVELS OF GOVERNMENT REGARDING MINIMUM COMPETENCY TESTING SHOULD BE CLEARLY DELINEATED. THE FEDERAL GOVERNMENT SHOULD NOT BECOME INVOLVED IN MINIMUM COMPETENCY TESTING IN STATES OR ESTABLISH A

Revised

NATIONAL TESTING PROGRAM OR NATIONAL MINIMUM COMPETENCY STANDARDS. TECHNICAL ASSISTANCE TO THE STATES, SUPPORT FOR RESEARCH AND DEVELOPMENT, AND DATA COLLECTION ON A SAMPLING BASIS AND REPORTING OF NATIONAL LEVELS OF KNOWLEDGE ARE IMPORTANT FUNCTIONS THAT THE FEDERAL GOVERNMENT SHOULD SUPPORT. GENERALLY, THE STATE SHOULD ESTABLISH THE BROAD GOALS AND OBJECTIVES FOR THE STATE, SET THE LIMITS WITHIN WHICH LOCAL DISTRICTS MUST OPERATE TO ACHIEVE THOSE GOALS AND OBJECTIVES, PROVIDE FUNDS AND TECHNICAL ASSISTANCE TO DISTRICTS IN PROGRAM DEVELOPMENT AND STAFF TRAINING, AND GIVE THE LOCAL DISTRICTS AND THE SCHOOLS LATITUDE TO DESIGN AND IMPLEMENT EDUCATIONAL PROGRAMS TO ACHIEVE THOSE GOALS AND OBJECTIVES. LIKEWISE, IF A STATE-WIDE MINIMUM COMPETENCY TESTING PROGRAM IS IMPLEMENTED, STATES SHOULD PROVIDE FOR LOCAL DISTRICTS TO HAVE AN IMPORTANT ROLE IN ESTABLISHING STANDARDS AND TESTS.

Important Considerations:

- The states, with constitutional sanction, have the primary responsibility for public education.
- A federal role in minimum competency testing and standards might include technical assistance to states, development of bench mark data on changes in achievement by population samples in important skills and competencies, and support for research and development of better methods for competency assessment.
- Standard setting is a shared responsibility by the state and local education agencies. A strong rationale exists for making many educational decisions at the district or local school level within each state. It will be important to determine the level at which decisions should be made about types of competencies to be

identified, desired level of achievement, amount and types of remedial assistance, types of data to be collected, and measures of achievement.

5. THE PROCEDURES AND PROCESSES FOR DEVELOPING AND IMPLEMENTING STANDARDS SHOULD ENCOURAGE A BROAD BASE OF ACTIVE INVOLVEMENT AMONG PARENTS, STUDENTS, EDUCATORS, AND OTHER CONCERNED CITIZENS.

FESSIONAL deleted

Important Considerations:

- Deciding what to include under basic skills, determining what the standards should be, developing good tests, and designing effective instruction programs are complex problems for which standard solutions do not exist. Because educators do not have all the answers, research directed toward the identification of better approaches to solving learning problems should be an integral part of the decision to test.
- If minimum competency testing and standards are to be implemented, sufficient time should be provided for planning and development work and for teachers and students to adjust to the new requirements. Components of the program which must be built into the time line include the selection of competencies, evaluation of the present curriculum and possible design of new instructional components, development and piloting of tests to establish reliability and validity, and the development of a diagnostic capability.
- While a wide range of useful skills have been suggested as appropriate competencies for schools to address, selection among basic or life skills, process and social skills, or skills and knowledge within each subject area can be made within the identified

jurisdiction. Because competency testing may or may not be used to measure student progress in all areas, it will be necessary to determine how large a role the minimum standards should have in the definition of a quality education.

- In selecting tests to be used, whether done at the state or local level, special attention should be given to determine that the tests measure what the curriculum is designed to deliver. This is often more difficult to do when tests are selected and mandated by the state, because of differences in instructional programs among school districts within a state.

Improved wording

6. IF MINIMUM COMPETENCY TESTING AND STANDARDS ARE TO BE USED AS A FACTOR IN PROMOTION OR GRADUATION DECISIONS, CAREFUL ATTENTION MUST NOT ONLY BE GIVEN TO THE CONTENT AND ADMINISTRATION OF THE TEST, BUT ALSO TO THE PROVISION OF COMPENSATORY AND BILINGUAL EDUCATION PROGRAMS, REMEDIATION, AND ANY OTHER INSTRUCTIONAL STRATEGIES NECESSARY SO THAT APPLICATION OF THE STANDARDS WILL NOT DISCRIMINATE AGAINST MINORITIES OR THE DISADVANTAGED.

Better transition ideas

Important Considerations:

- The demands for higher educational standards are bound to challenge many of the changes made in schools during the past decade to promote equality of opportunity. During that time, modifications in curriculum and standards were made in an effort to hold students in school. A decline in the dropout rate to an all time low of 25% in 1977 is evidence that the changes may have encouraged a larger number of previously disinterested students to remain in school.
- The College Entrance Examination Board panel found two thirds to three quarters of the decline in the SAT scores between 1963 and

1970 was related to the "notable extension and expansion of opportunity in the United States." Policymakers should study the impact that standards may have on students who may choose to drop out of school because they are unwilling or unable to meet the established standard.

- Minority groups are urging consideration of the impact of testing on culturally and linguistically different populations. In state-wide tests in some states, minority students have fallen below standard in large numbers. The needs of such students should be addressed without limiting access to opportunity through tracking or denial of the high school diploma. Alternative methodologies or environments will need development and support to meet the needs of students not achieving in the traditional setting.
- Other special need groups include special education students who may lack the capacity to pass a proficiency test, handicapped students who need special forms of the test, gifted students who deserve challenging work not provided by an emphasis on the minimum, and highly mobile students who must phase in and out of various local or state determined requirements.

7. DEVELOPMENT OF MINIMUM COMPETENCY TESTS AND ADMINISTRATION OF MINIMUM COMPETENCY STANDARDS COST MONEY. SUFFICIENT FUNDS SHOULD BE PROVIDED TO DO THE DEVELOPMENT AND IMPLEMENTATION WORK WELL. FUNDS ARE ALSO NECESSARY TO IMPLEMENT GOOD COMPENSATORY EDUCATION, BILINGUAL EDUCATION, AND REMEDIAL OR COMPENTENCY BASED EDUCATION PROGRAMS AND FOR STAFF TRAINING IF IMPROVED ACHIEVEMENT IN THE BASIC SKILLS FOR ALL PUPILS IS TO BE ATTAINED. THESE REQUIREMENTS WILL LIKELY CALL FOR REVISIONS IN

PRIORITIES FOR USE OF EXISTING FUNDS AS WELL AS PROVISION OF ADDITIONAL FUNDS.

Important Considerations:

- Funding decisions should be made as an integral component of minimum competency policy. Designing a testing program that directly affects students should not be undertaken without determining how the program, and the necessary remedial assistance to students, is to be financed. Some of the costs to be considered are:
  a. Administrative costs--record keeping and reporting expenses.
  b. Enforcement costs--staff to monitor, evaluate, and police.
  c. Costs in lost support to programs previously receiving funding.
  d. Test development costs.
  e. Test administration.
  f. Remedial work.
- Policymakers need to consider which are the most cost effective ways of assisting students who do not achieve at minimal levels. Because it will cost more to raise some students to the minimum than others, districts with large numbers of hard to educate students should be provided the necessary funding to implement remedial programs. Funding alternatives to consider include categorical grants based on number and type of low achievers, indexing the basic financing formula, and reprioritizing local budgets to create the necessary funds for remediation.
- Finances will necessarily determine how idealistic or realistic a state can be in setting standards. In the idealistic sense, "If

one believes that everyone can master most things given enough time (and time is money), then it follows that the more people who must master a skill, the more it will cost society."[4] For example, the resources needed to raise the reading level of a student from 2nd to 8th grade level will be significantly higher than for a single semester of remedial reading that is not aimed at a specific criterion. The resources and priorities of each state and locality should be evaluated when determining how high the standard should be.

8. MINIMUM COMPETENCY LEGISLATION AND/OR POLICIES SHOULD HAVE AS THEIR PRIMARY OBJECTIVE THE IMPROVEMENT OF PUPILS' MASTERY OVER THE BASIC SKILLS. SINCE UNIVERSITIES AND COLLEGES CAN MAKE A SIGNIFICANT CONTRIBUTION TO THE ATTAINMENT OF THIS OBJECTIVE, POLICY MAKERS SHOULD GIVE ATTENTION TO WAYS IN WHICH THE EXPERTISE LOCATED IN INSTITUTIONS OF HIGHER EDUCATION CAN BE BROUGHT TO BEAR IN IMPLEMENTING A MINIMUM COMPETENCY PROGRAM.

all new

Important Considerations:

- Teacher training institutions have a significant impact on the preparation and selection of persons who teach in the nation's schools. It is important for these institutions, through appropriate pre-service and in-service programs designed in cooperation with SEA's and LEA's, to insure that teachers, principals, and all other school personnel are adequately trained to carry out their responsibilities.
- The state of the art in the development and administration of tests leaves much to be desired. Universities and colleges should perform research and development functions in furtherance of knowledge and

[4]Barry Anderson, Costs of Minimal Competency Testing, a paper prepared for the ECS/NIE conferences on minimum competency testing.

practice concerning the technology of testing.

- If minimum competency standards and testing programs are to ultimately result in improvement of students' educational achievement, additional research focused at the teaching-learning process is required. Too little is known about the identification and treatment of specific learning disorders, potential remediation techniques and, more generally, what factors make educational programs effective. Until more is known about the learning process, our inability to teach some pupils the basic skills is likely to continue.

Pre-Measurement Issues in Minimal Competency Testing Programs[1]

Peter W. Airasian
Boston College

Introduction

Discussion about minimal competency testing programs generally focuses either upon specific implementation issues such as the pupil competencies desired, the nature of measurement techniques and setting standards of satisfactory performance, or upon the legal, ethical and measurement consequences of these issues. That such issues have received the bulk of attention is not surprising, since they encompass the practical problems faced by districts and states in trying to carry out mandated minimal competency programs.

But there is another, prior set of issues associated with minimal competency testing programs which also warrant attention. These are issues which arise in the earliest stages of program planning and they are important for three reasons: 1) they define the nature of minimal competency programs; 2) they impact heavily upon specific implementation decisions; and 3) they are at the root of measurement and legal challenges which arise in carrying out programs. The purpose of this address is to describe four pre-measurement issues: 1) the attempt to mandate or legislate pupil achievement; 2) instructional uncertainties; 3) the language of minimal competencies; and 4) the consideration given to the question, "Who is to be held accountable for learning?".

---

[1] Paper presented at the National Consortium on Testing, Spring Membership Conference, Arlington, Virginia, May 15-16, 1978. Portions of this paper were abstracted from P.W. Airasian, G. F. Madaus and J. J. Pedulla, "Policy Issues in Minimal Competency Testing and a Comparison of Implementation Models." Report submitted to the Policy Subcommittee of the Massachusetts Advisory Committee on High School Graduation Requirements, Massachusetts Department of Education, April 14, 1978.

It is appropriate to note here, for the purpose of sharpening the succeeding discussion, that although over 30 states have adopted some form of minimal competency testing program, no two states' programs are identical. While state programs differ on a number of dimensions, including the purpose of the program (diagnosis, assessment, certification for high school graduation, etc.), the level at which the program is carried out (statewide versus individual districts), and the time at which testing begins, my comments are directed towards minimal competency programs which tie measured pupil competence to the receipt of a high school diploma. This form of minimal competency testing is most controversial and best illustrates the importance of the various issues I will discuss.

1. The attempt to mandate or legislate pupil achievement.

The minimal competency testing movement is the latest response to a series of inter-related events occurring in education over the past fifteen years. Starting in the mid-1960's, large-scale investigations of school and compensatory education program effectiveness (e.g., Coleman et al., 1966; Cicirelli et al., 1969; Jencks et al., 1972) concluded that schools and school programs have very little direct, or unique, effect on pupil achievement. These studies implied that expenditures on small class sizes, teacher salaries and a host of other school resources had little impact on the achievement levels of pupils. More recently, the 14 year decline in national SAT scores has been discovered and interpreted by many as indicative of falling standards and a progressive erosion of pupils' basic skills. The 1972 "educational malpractice" suit filed by Peter Doe in San Francisco has raised questions about the meaning of a high school diploma and, faced with the threat of similar law suits, made educators more cognizant of the need

to certify pupils in basic skill competence. College teachers and employers have added their voices to the chorus decrying the lack of reading, writing and computation skills of high school graduates. These and other factors (Airasian, Madaus and Pedulla, 1978; Brodinsky, 1977) have led to the call for minimal competency tests to certify pupils for high school graduation.

What is most interesting in the rise of the minimal competency testing movement is the fact that the movement represents a reaction on the part of the general public to a perceived decline in the quality of education. A primary push for minimal competency programs has tended to come from the general public, through state legislatures and state boards of education. The direct result of this pressure has been a series of legislated or mandated competency programs which tie a pupil's test performance to high school graduation and which, in the end, are intended to raise pupils' achievement levels.

We must, however, be less than sanguine about efforts to legislate or mandate programs whose aim is to improve pupil achievement. Certainly the recent history of Head Start and Title I programs belies the notion that legislated programs, even those backed by a substantial investment of funds, necessarily will have a demonstrable impact on the achievement level of pupils.

We may contrast policy designed to affect pupil achievement directly with policy designed to affect pupil participation in or access to educational programs and resources (Wise, 1978). The latter problem, which may be considered a problem of equal opportunity associated with the distribution of resources and opportunities, is essentially a legal or political problem. When schools discriminate on the basis of race, handicap, or sex, action by the courts or higher authorities can redress the imbalance. Thus, goals associated with the distribution of resources or opportunities can be promoted through legislation and litigation. However, policy designed to solve

the problem of low achievement is different in kind from policy aimed at resource or opportunity distribution. The problems of low achievement are technical in nature, more related to the means of education than to its ends. For example, we may legislate equal access to schools for different racial and ethnic groups, but it is quite another matter to presume that equal access will, in and of itself, produce equal achievement for the groups. The latter aim requires a much more detailed consideration of the methods and strategies of instruction.

Similarly, for minimal competency programs, questions such as 'how does one define minimal competency?', 'can competencies be taught to all pupils with reasonable hope of success?', and 'are suitable and effective remedial strategies available to help hard-core underachievers attain the competencies?', must be addressed and resolved before we can hope to increase substantially pupil achievement. These are questions of means, not ends, and it seems clear that some legislatures and state boards of education have failed to recognize these as issues - or have assumed that they are easily resolvable - in mandating minimal competency programs.

The goals of eradicating illiteracy and of increasing pupils' achievements are surely appropriate educational goals. Whether these goals can be attained through legislated or other mandated efforts to install minimal competency testing programs in schools or through greater consideration and support of instructional research is a question to ponder. What is clear, however, is that the installation of minimal competency programs tied to high school graduation in the absence of consideration of means issues will pose problems at the implementation stage and thereafter.

2. Instructional uncertainties

Although the ultimate aim of minimal competency testing programs is to increase pupil achievement, most programs contain a series of intermediate aims

which define implicitly the process through which student achievement is to be increased. Thus, people variously see minimal competency testing as a device to make students work harder, to increase emphasis on the 3R's in school, to hold teachers and administrators accountable for poor pupil performance, to increase the cost-effectiveness of education, and to provide early diagnosis and remediation for pupils who are having learning difficulties. By and large, the lure of minimal competency testing programs resides in the perceived desirability of these aims, although historically it can be demonstrated that programs with similar purposes have not been overly successful in attaining such ends.

What is missing in the clamor for these ends is reference to the process of education - how educational practice affects or can affect the individual pupil. In essence, minimal competency programs, particularly those which tie test performance to the receipt of a high school diploma, are intended to alter the process of education, often without a clear conceptualization of how the process works or of its capabilities and limitations. Of course, the same statement can be made with equal validity for a host of other programs which have appeared in recent years.

But it must be recognized that minimal competency programs designed to certify high school graduation are unlike most other programs with which we are familiar. In most existing programs, be these statewide assessments, Title I projects, or local district achievement testing, inferences about student learning are made on the basis of aggregated data, summed and averaged over a large number of pupils. The consequences associated with individual pupil performance generally are not great or severe. Minimal competency testing programs in which pupil performance is tied to high school graduation do not focus primarily on the average performance of groups of pupils, but rather on the particular performance of

each individual. Moreover, the consequences of poor performance are grave for each pupil, namely denial of a high school diploma. Because the focus in minimal competency testing is on each individual pupil, and because there are severe penalties associated with poor performance, minimal competency programs pose a number of problems.

The focus upon individual pupil certification raises two general issues. The first is that improper instruction, faulty measurement procedures or inappropriate standards which may result in injury to individual pupils will form the basis for legal challenges once the program is underway. The second issue raised by focusing decisions upon the individual pupil relates to the characteristics of competencies selected and the instructional techniques available to teach these competencies. It is in considering this second issue that we are brought face to face with concerns about educational practices and processes.

Probably the single greatest concern in selecting competencies which will be tested to certify high school graduation is the answer to the question, "Is it reasonable to expect, given sufficient time and resources, that schools can foster these competencies in all children?". If students are to be held accountable for performance on the selected competencies and if there are sanctions associated with poor performance, it is critical that the competencies are viewed as capable of attainment by all pupils. It is unfair - and probably illegal - to hold pupils accountable for behaviors, skills and knowledge that are in part, or totally, beyond the school's direct power to control and influence. I recognise, of course, that it may be harder for some pupils to attain the competencies than others because of motivational, intellectual, physical, or emotional handicaps. Some pupils will undoubtedly require more time, more remediation, and more individualized instruction to attain competencies that their peers find all too easy

But if there is doubt about whether it is reasonable to expect all pupils to attain a competency as a result of instruction - of whatever type, duration, and time - then that competency should not be part of a program tied to high school graduation.

For certification in minimal competency programs, it may be insufficient to argue that "I can guarantee teaching, but not learning." A stronger case will need to be made that the competencies sought are available, at least potentially, to all students through the instructional process. Undoubtedly there will be pupils who will not master all, or even most, of the required competencies. There does, however, have to be a strong belief - or better yet, evidence - that the school could foster the competencies in all pupils given the correct set of circumstances. An important distinction must be kept in mind. There is a difference between the proportion of students one expects, under optimum conditions, could successfully learn the competencies and the proportion of students who actually attain measured minimal competency. Students may fail a minimal competency test for a number of reasons. Nonetheless, the expectation must be that given the right circumstances, each student could have mastered the competencies.

The preceding argument is important because in selecting competencies, schools implicitly go on record as assuming the major responsibility for fostering those competencies. If they cannot, it is unfair to expect pupils to manifest the behaviors and penalize them if they don't. This argument seems simple and straightforward, but it contains an important concern. The tendency to focus on the desirability of selected competencies diverts attention from the question of whether they are amenable to instruction in the schools. In fact, not a great deal of systematic, transferable knowledge exists about the instructional process. This

is particularly true for more generalized pupil behaviors such as the ability to apply ideas to new situations, or to synthesize disparate information into a coherent whole, not to mention a wide range of social and affective behaviors.

Some have made a distinction between two kinds of pupil competencies. The first type of competencies are called 'enablers', school-specific skills such as multiplication facts, spelling, punctuation rules, etc. The second type are more generalized 'competencies'; those involving the application of school-specific skills and knowledge to more life-like situations such as balancing a checkbook, computing compound interest, etc. The state of the art in instruction is such that we are considerably more confident about teaching pupils 'enablers', basic facts and knowledge, than we are in effecting the transfer of these skills to life situations.

While most educators would agree that basic cognitive skills and processes do fall within the domain of teachable behaviors, there appears to be less certainty that so-called 'higher level skills' such as the ability to apply knowledge to the solution of unfamiliar problems, the ability to recognize unstated assumptions, or the ability to formulate appropriate hypotheses based on an analysis of the factors involved in an unfamiliar situation, are capable of being successfully taught to all pupils. Certainly the relationship of general intelligence to mastery of these more abstract behaviors belies the notion that they are readily available to all students (Madaus, Nuttall and Wood, 1973). Moreover, and perhaps most crucial for a competency-based certification approach, it is not at all clear how pupils acquire such behaviors or what types of instructional materials are most appropriate for developing them. The key issue here is not that these behaviors cannot be taught, but rather that teaching them, given the present state

of knowledge, is as much an art as a science. Techniques useful with some pupils are not useful with others; strategies adopted by one teacher are not readily exportable to other teachers. Certainly, minimal competency programs which seek such outcomes will face instructional problems which involve a degree of complexity far removed from the simple 'diagnose and prescribe' language contained in the description of most competency-based programs (Madaus and Airasian, 1977).

If the domains of personal, social and career development are to be tested as part of a minimal competency program, and if these domains involve more than cognitive and psychomotor competencies, then it may be particularly difficult to design appropriate instructional techniques for the development of such competencies. Even when the issue of whether the school should teach confidence, self concept, work ethic, or job preparedness are put aside, there is a serious question about whether schools can teach such affective competencies to all pupils. To date, programs in these areas have not met with marked success, and the prospects for widespread, universally successful teaching of such objectives remains a hope rather than a reality. An additional set of legal problems may be associated with efforts to teach and evaluate mandatory personal and social behaviors. Note that all of the preceding concerns are relevant not only in the context of regular school instruction, but also in the context of efforts to remediate pupils' learning difficulties.

Focusing upon the ends of programs rather than upon the means available to attain these ends has particularly important overtones in minimal competency programs. In such programs an understanding of the limits of the state of the art in instruction and curriculum is critical, because by the explicit publication of minimal competencies, schools go on record as implying that they can in fact, develop these competencies; schools imply that they can foster the

competencies in all pupils. There is nothing wrong with wanting each pupil who graduates from high school to manifest competencies which will help him cope with life in a complex technological society. However, given present day knowledge about instruction, we should think carefully about translating our desires into a system of accountability which places the primary onus of responsibility on the individual pupil.

3. The language of minimal competence.

A third issue which is related to and tends to exacerbate the problems already noted concerns the language of competency. I noted above that the language of competency tends to be ends oriented, focusing upon the potential products of programs rather than upon the means necessary to attain these ends. An emphasis on ends or outcomes is not unexpected, since much of the rhetoric about minimal competency programs has been generated by advocates seeking to propagate their belief in the efficacy of such programs. Moreover, much of the appeal of minimal competency is a function of its stated ends rather than a consequence of any evidence of the ability of such programs to attain these ends. The language employed in discussions of minimal competency programs unfortunately has the potential to raise expectations about the ability of schools to make all segments of the student population literate, without the certainty that schools can actually do this.

Consider, for example, the language used by the California State Legislature when it passed a law requiring junior and senior high schools to adopt standards of proficiency in basic skills and tied high school graduation to these standards as of 1980: "...pupils attending public schools in California acquire the knowledge, skills, and confidence required to function effectively

in contemporary society." Such language is not unique to the California Legislature and similar themes can be found in bills of other states. At first blush, it is difficult to fault such an intent. It holds out the promise of knowledgeable, confident pupils, successfully carrying out their chosen roles in society. Few would debate the merits of such a goal.

However, if we consider more fully the words used in the intent we shall see that there are a number of questions left unanswered by the legislature's statement. These questions are important since in two years receipt of a high school diploma in California will depend upon their answers. Consider the wording more closely. What, for example, is the distinction between knowledge and skills? Is knowledge a familiarity with basic facts such as the multiplication tables, rules of punctuation, etc., while skills are applications of these facts to real life situations such as filling out a tax form, writing a coherent business letter, etc.?" And if this distinction is correct, are we confident that schools can - or should - teach such applications to all pupils? Should every pupil have to be able to fill out a tax form, compute compound interest, etc., in order to receive a high school diploma?

More to the point, however, are the questions raised by the words used in the latter portion of the statement. Can - or should - schools teach confidence? What exactly is confidence? How much confidence is too much and how much is too little? Is the possibility that some pupils may fail to demonstrate the cognitive proficiencies required for obtaining a high school diploma antithetical to the goal of instilling confidence in all pupils? What of the pupil who can demonstrate cognitive competencies but lacks confidence? Should he/she be denied a diploma? At a more general level, the intent to instill an affective behavior such as 'confidence' in all pupils raises a number of legal, ethical and political questions about the aim and function of schools in the affective domain.

Finally, what skills, knowledge, etc., are 'required' to 'function effectively' in 'contemporary society'? There is a legitimate question - and this question should be examined empirically - as to the number of American adults who presently possess the basic skills and knowledge talked about in many minimal competency programs. What then, does 'function effectively' mean? Is there a universal norm of effective functioning which can be applied to all people, regardless of occupation, geographic location, level of education, etc., to define the ability to function effectively? Moreover, in a society which is changing constantly, is it sufficient to set our goals at effective functioning in 'contemporary society?' 'Contemporary society', by definition, will not be contemporary in five or ten years and the skills, knowledge and confidence required to function in 1978 may be quite obsolete by 1988.

I have singled out the California legislation for discussion, but I might have made the same argument using other states as examples. Had I done so, the arguments made would have been similar, but the terms would have been different (e.g., survival skills, minimal competence, life skills, functioning members of society, etc.).

If one examines the California State Legislature's intent vis a vis its minimal competency program, one sees that the second half of the intent,'required to function effectively in contemporary society,' is really a justification for the first half, 'acquire the knowledge, skills and confidence.' We have just seen that the fuzziness of this terminology and the questions it raises cloud both the intent and specifics of the legislation. It appears that there are at least two ways to justify the need for a minimal competency program. The first is to justify the program in terms of some wider societal functioning and adequacy, which is what the California legislators have done. The second

is to say simply that high school graduates should have certain competencies because the community believes and wants pupils to have them, thereby avoiding statements which directly tie competence to 'survival' or 'effective functioning' in society. The way in which a legislature or state board of education chooses to justify its minimal competency program directly impacts on the type of validity a test of minimal competency must possess. The validity question, in turn, inevitably raises other legal and ethical problems.

It needs to be understood that the rationale for the adoption of any minimal competency program resides in the value domain. Whether or not a state or district has a minimal competency program will not be decided on the basis of empirical evidence. Such evidence may explain or describe particular characteristics of the program, but the evidence itself will not determine the program's adoption or defeat. People will do this, and they will do it on the basis of value judgments about the worth of the program and the competencies such a program encompasses.

I do not wish to prolong unduly this discussion about the language of competency, but it is worthwhile to point out the seductive nature of the terminology used to describe and justify minimal competency programs. Undefined, perhaps undefinable terms are used in discussing such programs, and it is only when one thinks through the meaning and application of such terms that the apparent simplicity of minimal competency testing is stripped away, revealing its true complexity. The present language of minimal competency programs often does a grave disservice by masking this complexity and by falsely raising expectations about what programs will accomplish.

4. The consideration given to the question, "Who is to be held accountable for learning?"

Noted previously was the fact that different publics hold different educational groups responsible for students' poor achievement. Some hold students themselves to blame, others fault teachers who are ignoring "basics" to focus on creative or humanistic objectives, still others hold administrators responsible, and some simply place the blame on the "system" itself. Although different groups wish to hold different participants in the educational process accountable for student achievement, it is clear that when minimal competency testing is tied to high school graduation, the onus of accountability falls squarely on the individual pupil.

There are many reasons for this, ranging from the failure of states and locales to consider in advance of program implementation just who they wish to hold accountable, to political and logistical problems in trying to hold anyone save students accountable for their own learning. Given that much of education is sequential and cumulative and that most skills measured by minimal competency tests are learned over a number of grades, it is very difficult to hold an individual teacher accountable for the fact that at grade 8 or 10, a pupil cannot demonstrate mastery of certain basic skills. While one could quite justifiably argue that the system is at fault, the only individual who can be held directly accountable is the pupil. So long as minimal competency programs restrict the assessment of competencies to the junior or senior high school years--after two thirds of a pupil's time in school is completed--it will be extremely difficult to hold any individual teacher or school accountable for a pupil's performance on a minimal competency test.

It is not my purpose to argue here that students should not be held accountable for their learning. Rather, I am trying to point out that by default,

largely as a result of the procedures adopted in some states, the student assumes the primary onus of accountability. Since the potential consequence of failing a minimal competency test is grave--denial of a high school diploma--and since there are many unresolved issues associated with the capability and limits of school instruction, it seems important that the question of accountability be explicitly considered before a minimal competency program is enacted.

It is important to ask, "Whom do I wish to be accountable, if anyone?". Is the concern primarily with teachers, pupils, or school curricula in general? If sanctions for student failure are to be applied in a minimal competency program, it is necessary that the reasons for applying the sanctions to a particular group be clearly thought through in advance of program implementation. We may ultimately decide that the student bears final responsibility for his learning, but we should decide this explicitly, considering all that I have noted previously, not let it be decided by default or by the particular way we implement a minimal competency program, as has happened in many states.

The question of accountability obviously is not a simple one. It involves the decision to hold some group or the 'system' responsible for student learning and hence has political, economic, educational and ethical aspects. I have not argued that any particular individual, group or organization be singled out as the object of accountability. I have suggested instead, that since a minimal competency program is a form of accountability, a potentially harsh form, it is essential that the issue of who is to be accountable, and how accountability will be determined, be thought through before the program is instituted.

In thinking through the problems of accountability it may become apparent that primary concern is with hard-core underachievers rather than with all students who attend high school, or with a few teachers who ignore 'basics' rather than with

teachers in general. Moreover, it may be recognized that competency testing at grade 8 or above is inefficient for identifying pupil learning needs at a time when remediation can be most beneficial. Consideration of such issues prior to program definition and implementation not only could alter substantially the nature of a minimal competency program initially envisioned, but it also could overcome many of the measurement and legal concerns which will inevitably ensue from failure to consider these issues in the early stages of thinking about minimal competency testing.

Conclusion

I have tried to highlight four pre-implementation issues and concerns which should be addressed before installing minimal competency testing programs. I have suggested that these areas have not received adequate attention in many of the minimal competency programs which have been implemented to date. The problems of minimal competency testing programs are not confined exclusively to questions about testing procedures and standard setting. Rather, the spectre of minimal competency gives rise to more fundamental questions associated with the instructional capabilities and limitations of schools, the propagation of false or inflated expectations, and the appropriate locus of responsibility for pupil learning. While such questions clearly are related to the search for suitable standards and measurement techniques, they also are important in their own right and should not be overlooked.

References

Airasian, P.W., Madaus, G.F., and Pedulla, J.H. Policy issues in minimal competency testing and a comparison of implementation models. Report submitted to the Policy Subcommittee of the Massachusetts Advisory Committee on High School Graduation Requirements, Massachusetts State Department of Education, April 14, 1978.

Brodinsky, B. Back to the basics: The movement and its meaning. Phi Delta Kappan, 58, 7, 1977, 522-527.

Cicirelli, V.G., et al. The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development. Volume 1, Westinghouse Learning Corporation and Ohio University. U.S. Department of Commerce/National Bureau of Standards/Institute for Applied Technology. Distributed by Clearinghouse, Springfield, Va., PB 184 328, 12 June 1969.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D. and York, R.L. Equality of educational opportunity. Washington, D. C.: Office of Education, U.S. Department of Health, Education and Welfare, 1966.

Jencks, C., Smith, M., Acland, H., Bane, M.J., Cohen, D., Gintis, H., Heyns, B. and Michelson, S. Inequality: A reassessment of the effect of family and schooling in America. N.Y.: Basic Books, 1972.

Madaus, G.F. and Airasian, P.W. Issues in evaluating student outcomes in competency based graduation programs. Journal of Research and Development in Education, 10, 3, 1977, 79-91.

Madaus, G.F., Nuttall, R.L. & Wood, E. A causal model analysis of Bloom's taxonomy. American Educational Research Journal, Vol. 10, No. 4, 1973, 253-262.

Wise, A. E. Minimum competency testing: Another case of hyper-rationalization. Phi Delta Kappan, 59, 9, 1978, 596-598.

Measurement Issues and Consequences Associated with Minimal Competency Testing[1]

George F. Madaus
Boston College

While minimal competency testing is one of the strongest movements in contemporary education, testing programs designed to certify a pupil's successful attainment of a given level or standard of education are not new. Quite the contrary, such programs have a long history. The first such program in the United States began in 1845 with the annual written examination of the 'first' class of the Boston Grammar Schools. In 1877 the New York legislature empowered the Regents to establish a system of exams "to furnish a suitable standard of secondary school graduation....and admission to the several colleges of the State" (Spaulding, 1938). The British system has for a long time used an external examination system to certify the successful completion of various levels of secondary education.

Such certification programs have not been limited to the secondary level. In the latter half of the 19th century, in England, Ireland and Jamaica, pupil results on exams of reading, writing and arithmetic determined all, or a portion of, the teacher salary. In 1927, the Irish Government instituted written exams in Irish, English and arithmetic to certify a pupil's successful completion of primary school. There is a contemporary ring to the 1927 argument advanced in Ireland that the primary certificate exams would give teachers a "clearer knowledge of the standards

---

[1]Paper presented at the National Consortium on Testing Spring Membership Conference, Arlington, Virginia, May 15-16, 1978. Portions of this paper were abstracted from P. W. Airasian, G. F. Madaus and J.J. Pedulla, "Policy Issues in Minimal Competency Testing and a Comparison of Implementation Models." Report submitted to the Policy Subcommittee of the Massachusetts Advisory Committee on High School Graduation Requirements, Massachusetts Department of Education, April 14,1978.

after which they are expected to strive; the formal character of the examination would be a powerful stimulus to pupils. The certificate would come to be recognized as qualifying for entrance [to secondary level education] and would have considerable value for the purposes of employment" (Ireland, 1954). However the Primary Certificate Program never lived up to these expectations and was abolished in the early sixties.

In this paper I would like to discuss briefly some of the major consequences for pupils, teachers and the curriculum associated with testing programs designed to certify the successful completion of a given level of education that I feel have implications for present day minimal competency testing programs. Further, I will outline specific technical issues that inevitably arise with the institution of minimal competency testing programs.

Possible Consequences Associated with Minimal Competency Testing Programs

Make no mistake, the certification of competencies for graduation has the potential to re-orient the aims and emphases of education. Examinations which are used to make important decisions as major disjunctions in the educational system have great effects on teachers, pupils and the curriculum (Bloom, 1969). Past experiences with certifying exams clearly show that the degree to which the system is changed is a direct function of the degree to which teachers, pupils and parents perceive the test as a real obstacle to certification.

Driving exams, which are minimal competency tests, are taken seriously by most people because the license to drive is highly prized and further because there is a significant first-time failure rate. Everyone knows someone who failed his/her driving test and partially because of this people spend considerable time, effort and money in preparation.

Once the "credibility" of the certifying exam has been established---that is,the sanctions associated with the competency exams are widely perceived as real-- then the consequences for education will in large measure be determined by what Europeans call "the tradition of past examinations." This tradition begins to determine what is taught, how it is taught, what pupils study and how they study.

A survey of external exams like those mentioned above shows that the tradition of past examinations is a double-edged sword. On the positive side, well-defined and valid performance measures have been powerful forces for re-directing teaching to cover better neglected areas and competencies. Vital curricular changes are often brought to fruition by altering the certifying examination to reflect the new objectives (cf. Morris, circa, 1969; Commission on Mathematics, 1959; Bloom, 1950). On the negative side, areas of the curriculum not directly examined for certification were often de-emphasized. In fact, when the external examination is a powerful determiner of future educational or life chances, teachers could do pupils a disservice by emphasizing new or non-examined objectives or competencies. Innovative teaching or the teaching of objectives not tapped by the certifying examination may actually hamper pupils' life chances by taking time and emphasis away from the objectives society utilizes to appraise competence. After a period of years many teachers become skilled in coaching pupils for the certifying measures and resist change in the system whose rules for success they have mastered.

Faced with a choice between one set of objectives which are explicit in the course outline and a different set which are explicit in past certifying examinations, students and teachers generally choose to focus upon the latter. This finding holds true over different countries and over many decades (Madaus and Macnamara, 1970; Spaulding, 1938; Norwood Report, 1943; Morris, circa 1969.

Also see Mukerji, 1966; Hotyat, 1958; Bloom, 1961; Tyler, 1963; Davies, 1965 Koerner, 1968; Srinivasan, 1971). Most studies have found that the proportion of instructional time spent on various objectives was seldom higher than the predicted likelihood of their occurrence on the external examination.

The impact of present day minimal competency testing programs like other certification exams, can be double-edged. The hope is that such programs can leadto the inclusion of, or increased emphasis in the school curriculum on skills and competencies previously neglected or treated only informally. Further, there is the potential to alert students and teachers to the specific competencies valued for graduation and thereby focusing teaching and learning upon these competencies. There is the likelihood that the general public, including parents and employers, will be better aware of the competencies implied by a high school diploma. These and a series of other beneficial outcomes are all potentially available as a result of the addition of competency certification as a requisite for high school graduation. Undoubtedly these reasons represent a large part of the justification for such programs.

These potential advantages should not divert attention from the potential disadvantages. As we saw, past experience with external exams shows that there is the clear danger that teaching and learning can become narrower, focusing upon competencies to be evaluated to the point of cramming or neglecting important outcomes not covered in the minimal competencies. The importance of course credits and content can thereby diminish in the eyes of teachers, pupils, and parents. As we saw with external exams, teaching and learning can become mechanized and less spontaneous, devoted to preparing for the competency evaluation rather than to educating pupils in the more general sense of the word. Most importantly, schools within schools may begin to emerge, with pupils grouped in terms of whether they represent the mass of students likely to be certified as possessing minimal competence or whether they represent those who are at risk in terms of being certified.

Unfortunately past experience tends to show that the detrimental edge of the exam blade is sharper and cuts deeper than its beneficial edge. Holmes (1911) long ago recognized the most important danger of external exams when writing of his experiences as a Chief Inspector in the Irish system. He observed:

> Whenever the outward standard of reality (examination results) has established itself at the expense of the inward, the ease with which worth (or what passes for such) can be measured is ever tending to become in itself the chief, if not sole, measure of worth. And in proportion, as we tend to value the results of education for their measurableness, so we tend to undervalue and at last to ignore those results which are too intrinsically valued to be measured (p. 128).

In summary, past experience with external certifying exams leads us to expect that minimal competency certifications will impact for better or worse on teachers, pupils, the curriculum and school organization.

Technical Issues Associated with Measurement of Minimal Competencies

Leaving the more general consequences associated with certifying competencies for graduation, a host of technical issues associated with the measurement of minimal competencies arise.

1. Setting standards for passing a minimal competency test

We saw earlier that certification exams are taken seriously when there are serious consequences associated with passing or failing and further when it is clear that not everyone passes the exam. Thus standard setting is a key element in any certification testing program. Unfortunately, the process is largely subjective and judgmental in nature. While much has been written about statistical approaches to determining the particular score that will minimize wrong decisions, they all ultimately rely on the ability to make subjective judgments regarding the relative loss incurred in erring in one direction as opposed

to the other. Comparisons of alternative statistical approaches, all technically sound, yield very different results (Glass, 1977). Thus, although at first glance it appears as if the statistical technology were available for determining cut-scores, there is no consensus on any one approach, and all approaches rely on subjective judgments of some sort.

In reality, the determination of a cut-score is largely political. "What percentage of students can we reasonably fail?" In considering this question officials can find themselves in a dilemma. There are obvious monetary, educational and political costs involved in failing a high percentage of students. Florida's experience with their math test bears testimony to this fact. On the other hand, if practically no one fails the test, the test can become meaningless, and authorities will be criticized for having set standards too low.

2. Issues of norm-referenced and criterion referenced testing

A norm-referenced test is one in which the scores are interpreted in terms of the pupil's performance relative to that of a comparison group. Norm-referenced scores generally take the form of grade equivalents, percentiles, or standard scores. Some minimal competency testing programs use a commercially available, standardized, norm-referenced test as their test of minimal competency. The average performance of students at a particular grade level then becomes the standard accepted as indicative that the student has mastered the minimal competencies. For example, a student might have to perform at the level equivalent to the average performance of students beginning the ninth grade. In other words, a grade equivalent score of at least 9.0 becomes the score a pupil must obtain to be certified.

There are a number of problems associated with this practice. First, the commercially available norm-referenced test de facto defines and selects the

competencies being certified. It may be extremely difficult to find an available test that covers the skills which the district or state wishes to certify. Further, the question inevitably arises, why accept the average performance of a ninth grader as the criterion for a high school diploma, why not the performance of the average 12th grader? There is a very good reason for this. A grade equivalent of 12.0 would correspond to the median score attained by the 12th graders in the norming population. Consequently one-half of the 12th graders fall above this point and one-half below. To set the cut score at 12.0 therefore would mean that only one-half of the 12th graders and even fewer juniors and sophomores would be certified. Thus a grade equivalent of 9.0 is politically more acceptable. To circumvent the problems of using a grade equivalent score as the criterion for passing a minimal competency test, it may be wise to avoid the language of grade equivalents altogether.

If a norm-referenced test can be found that encompasses the desired competencies or if one is built from scratch, then standard scores might be a preferable metric to use in setting passing scores. Percentiles do not have the problems of grade equivalents but could be confused by the public with percentages used in traditional grading or marking.

Now let us turn to criterion-referenced tests which do not reference a student's performance to that of a norm group. Instead, there are two meanings associated with criterion-referencing (Glass, 1976). The first is that a standard of performance is predetermined, and students meeting or exceeding that standard pass, or are classified as having mastered the competencies, whereas students who score below the predetermined cut-off score fail. The intended purpose of a criterion-referenced minimal competency test(s) is to separate students into two groups, those who have mastered the competencies and those who have not. This type of criterion referencing is not without its problems, however. The major problem being the one pointed out above, that is, how does one determine in advance what an appropriate standard or cut score should be?

The second meaning of criterion-referencing is that the test is designed and constructed in a manner that defines explicit rules linking patterns of test performance to behavioral referents (Jackson, 1970). Some people mistakenly think that a score of 80% on a minimal competency test gives us information about what a pupil can do in some absolute sense. This assumption is not completely correct. It is true that the percent of items answered correctly by a pupil is independent of how his/her classmates do on the test and hence is not norm-referenced. However, unless all the items on the test are clearly linked to a group of related competencies, the percentage score tells us nothing about what a student can or cannot do relative to the related competencies in any absolute sense. A score of 80% can be said to be criterion-referenced only in the sense that some arbitrary criterion or cut score has been set. It is not criterion-referenced in the second and more widely accepted sense. This point needs to be kept in mind when publicity about minimal competency tests start to be formulated by the sponsoring agency. To date criterion-referenced tests referenced to a skill domain have not been employed as part of minimal competency graduation programs. It is misleading to bill the competency tests as criterion referenced unless there is a link to a clearly specified skill domain.

Interestingly, if each district is permitted to set its own certifying test and standards then the comparability of the diplomas issued by different districts is questionable. A standard set at 70% of the items answered correctly by one district on their test may not be equivalent to the 70% set by another district on its test. Even if the tests are measuring the same competencies, the item difficulty levels could vary, thus making one test considerably harder than the other. I will return to issues of item difficulty presently. In the British system where the tests are uniform within a year but different from year to year, it is recognized that obtaining a pass or honor score one year is easier than in another. Comparability of standards between districts and over time is an important consideration that must not be overlooked or lightly dismissed.

3. Total vs. part scores

Related to the problem of standard setting is the issue of the nature of the score used to certify the competencies. A single total test score, either across subject areas or within a single subject area, may be suitable for making a pass/fail decision for certification purposes, but is virtually useless for the purposes of diagnosis and remediation. The total test score approach allows for the possibility of a student attaining the cut off, or criterion score, by correctly answering all items related to certain competencies, while incorrectly answering all items related to other competencies. If this happens, should that student be certified as having attained the required competencies? Further, if a test covers heterogeneous competencies, and a single cut score is used to determine pass/fail, then questions about the meaning attached to that score can be raised. In such a situation the agency would not be certifying each competency but instead that the pupil answered a certain percentage of heterogeneous items correctly. Such a procedure would hide more than it shows and can easily mislead the public who might naively believe that the separate competencies are being certified.

If the state or district wished to certify each competency identified by panels as desirable, then a separate test for each competency would have to be built. This would be fiscally expensive and could quickly become prohibitive in terms of the testing time required of candidates. A competency by competency scoring procedure also increases the record keeping task.

An alternative would be to build a separate test for each major class of competencies (i.e., computation, reading, writing, listening, etc.). The drawback to this approach is that a pupil could pass two or three tests and fail the remainder. As in the case of the external exams in England and Ireland, the certifying agency might have to certify performance on each test but require that all tests be passed before a diploma was issued, thus complicating record keeping.

Separate scores for separate competencies would give more information in terms of performance. However, it should be pointed out that even if separate scores are used, their diagnostic value in terms of planning remediation programs for an individual or group is slight. A separate diagnostic test would have to be built to pinpoint specific weaknesses. The construction of diagnostic tests involves developmental validation costs over and above those associated with the construction of the certifying exam.

There are no clear answers to the issue of one total score versus scores for each competency. The better approach should stem from the purpose of the minimal competency testing program. If the purpose is simply to categorize each student into either the pass or fail group without regard to diagnosis and remediation for those who fail, and without regard to whether specific competencies were mastered, the total score approach may be defensible. If the purpose is to do more than that, the competency by competency approach is more defensible.

4. Issues of test security

One problem that must be anticipated in setting up a minimal competency testing program is that of test security. Completely secure testing programs in which the same test, or large portions of it, are used every year are fairly common. However, in minimal competency programs it may not be possible or desirable to build completely secure tests.

External testing programs tied to high school certification historically have been secure only to the point of administration. In past years, old New York State Regents exams could be purchased in bookstores throughout the State. In the British system, the exams move into the public domain once they have been administered. In fact the tradition of past exams in external testing programs is inferred from available old exams, and, as we saw, dictates how pupils study, what they study and how teachers teach.

The situation of a test linked to a high school diploma is not comparable to those programs run by professional testing and certification agencies. In the latter case, the examinee voluntarily elects to sit for the exam. The examining body provides a service to the examinee for a fee, part of which goes to defray the expenses involved in building a test, running the program, and reporting results to the examinee or to schools designated by the examinee. In the case of certifying boards in medicine, nursing, or other professions, the applicant is applying for membership in a private organization. The applicant pays a fee to sit for the organization's certification exams which once again goes in part toward test development, maintenance and program costs.

The minimal competency test program is quite another matter entirely. First, secondary pupils do not elect, but for all intents and purposes are required, to sit for the test(s). Second, the test(s) is paid for out of tax funds. If a pupil fails the exam and is denied a diploma until he/she can pass it, it is quite conceivable that the applicant will want to question parts of the exam. Under the freedom of information law the student might be able to demand to see the test. If not, then certainly mini-Harrington Bills will be passed. Once the taxpayer's monies are involved, it is only a matter of time before someone asks for the test's release and the State will probably have to accede to such a request. If this analysis is correct then measures will have to be taken to make each year's test comparable and as far as possible make the standard for passing comparable from year to year.

Another reason for making the tests public after use is that their availability permits the public to judge how well the tests measure the things it thinks are important and should be certified before a diploma is issued. Further, Jencks (1977) points out that teachers who do see present commercial achievement tests argue that the tests don't measure things they want students to do or to know about.

Some of the testing profession cavalierly dismiss this reaction as teacher defensiveness. However, until tests are available for public scrutiny such a charge of defensiveness cannot be intelligently countered. The same argument holds for minimal competency tests.

Another issue under test security is that of cheating. Techniques for the detection of cheating will have to be considered. Techniques to assure that the person sitting for the test is in fact the person to be certified would have to be developed. This would best be handled by the local education authorities. It should be noted that if local teachers are involved in administration it would be next to impossible to keep a test completely secure over the long term. This does not mean that groups of teachers cannot be involved yearly in test construction; in England teachers play an essential role in developing the external exams (Brimer, Madaus, Chapman, Kellaghan and Wood, 1978).

5. Issues of item difficulty

In building a minimal competency test, it becomes imperative that items are tried out beforehand to get estimates of their difficulty before the test is finalized. The reason for this is simple; as is the case with traditional norm-referenced tests, the items are the individual building blocks that eventually produce the desired score distribution. Average item difficulty ultimately impacts upon the percentage of people who are certified. Thus, item difficulty cannot be separated from political considerations in setting the cut-score. Glass (1978) has argued quite convincingly that the reason why 36% of the students in Florida failed the math portion of the minimal competency test while only 8% failed the communication skills portion is that the average item difficulty level of the former was probably considerably lower than on the latter. He further points out that someone writing items to cover a given topic can produce some that turn out

too hard and others that turn out too easy but which on inspection appear to be equivalent. This phenomenon has also been encountered by the National Assessment (Greenbaum, et al, 1971). It is important to remember that the inference made by Florida officials was that considerable money needed to be spent on remediation in math, not that the math items were more difficult than those used to measure communication skills.

There are other reasons for pre-testing items. Item analysis can reveal ambiguities in the item stems or distractors. Further, it can reveal items which are differentially difficult for different groups of pupils. If the test makers believe that the level of achievement should not differ between two groups (say boys and girls), they can reject items which are more difficult for one or the other of the groups. This procedure would result in a test in which members of both groups perform equally well on average.

6. Issues of test validity

A central measurement issue in any minimal competency test program is the validity of the certifying test(s). While I shall not go into a detailed technical discussion of validity, certain issues concerning the validity of minimal competency tests need to be mentioned. If the minimal competency tests are used to certify that the pupil has acquired certain basic scholastic skills in areas like computation, reading, writing, speaking, listening, etc., then the test(s) must be shown to be content valid. To demonstrate content validity it is first necessary to define or specify the universe or domain of interest (i.e., computation), then describe or specify the skills and/or behaviors that are part of the domain (i.e., adding two digit numbers that involve carrying, etc.). Once

the domain is specified, content validity depends on the ability to demonstrate that the test item in fact represents the domain. If a test is content valid, then inferences can be made about a pupil's performance on the domain. Two facets of content validity appear in the minimal competency literature, curricular and instructional validity (McClung, 1977).

Curricular validity involves demonstrating that the items represent content and behaviors included in the curriculum materials used in the schools. This curricular facet of content validity is important because if the competencies are not explicitly reflected in the curriculum, then presumably they must be acquired through extra school experiences, and therefore it is unfair to include them as part of the high school graduation requirements. If the domain of interest is defined in terms of school specific skills, then unless the test is set by a local district, it would have to measure competencies common to all of the curriculum material used throughout the state for the subject in question. When the test is not set by the local district, demonstrating curricular validity can become complicated; the certification test is given between grades 9 and 11, and a host of curricular materials have been employed by various districts across a state over the grades prior to the grade at which competencies are being certified.

Instructional validity is a concept introduced by McClung (1977) who argues that it is not enough to demonstrate that the competencies are common to the various curricula; it must also be shown that an individual pupil received instruction directly related to the competencies required for certification. Such a demonstration could involve rather extensive record keeping. For example, teachers at all grades might be required to keep a log of the time spent teaching each competency for each pupil. The homework assignments, classroom tests, etc., for each pupil might also be used as indicators of "instructional validity."

Edmonds (1977) has argued that instructional validity is closely linked to the legal concept of 'right to treatment'. The courts have recently ruled that the presence of adequate treatment is a necessary condition before a person can be involuntarily confined to a mental hospital. Using an analogous argument, the court might rule that before an agency could deny a student a high school diploma because he/she lacked some set of minimal competencies, the agency must show that the pupil - who was compelled under law to attend school until he/she reached age 16 - received 'adequate' treatment. If such a decision were ever made, then demonstrating the 'instructional' validity of a certifying test would become very important.

If the competencies are worded in terms of life skills, i.e., balancing a checkbook, filling out a job application - determination of content, curricular and instructional validity is complicated. These life skills essentially ask the pupil to apply basic school skills to new situations. As we saw, the hope is that specific school skills like addition, reading, etc., in fact transfer to more general life skills like balancing a checkbook or reading a prescription label. Content/curricular validity may be quite hard to demonstrate for a test measuring life skills for two reasons. First, it may not be possible to define and delimit the domain of life skills. Second, such skills are not always common to the curricular materials used in the schools. Interest rates, checkbooks, tax forms, etc., were included in the 7-8 grade arithmetic books in the 40's and 50's but dropped out of sight in the 60's with the advent of modern math, the exception being courses like business arithmetic where such topics are still covered. Further, unless special courses, or blocks of time were specifically devoted to teaching the application of basic skills to life situations, it might be impossible to demonstrate instructional validity.

When the number of individuals to be certified is large, then direct measurement is often costly and time consuming, and indirect measures are often substituted. However, when the competencies are stated in terms of life skills, where the student is required to apply basic skills of literacy and numeracy to real life situations the construct validity of an indirect paper and pencil multiple choice measure of many of these skills becomes an issue. Many "life skill" competencies are validly measured by the most direct means possible--situational or performance tests — to determine if the student can or cannot perform these activities. Again, driving a car is a good example. Even when the competencies are limited to traditional scholastic areas such as writing or listening, the validity of using indirect multiple choice tests to assess these skills is questionable. To frame multiple choice items that ask a student to identify an example of poor, or incorrect writing, is no substitute for the more expensive and direct measure of asking the student to write a paragraph or a business letter. In the case of writing and listening, the relationship between such direct and indirect measures is very low and therefore the indirect measure should not be used. Any indirect, or surrogate measure will have to be validated against direct performance measures before it can be used. If the relationship between the direct and indirect measures is low then the direct measure must be employed.

So far I have avoided the issue of measuring affective competencies such as the development of confidence or interest in reading or such vague competencies as becoming a "life long learner." At the risk of being glib, these so-called competencies should not be part of a graduation requirement. Ethical concerns apart — and I don't mean to dismiss these lightly — relative to our sophistication in evaluating cognitive outcomes of learning, our skill at evaluating these types of competencies is small.

Indirect methods such as using a questionnaire to assess behavior are sufficient if one is interested in group performance data and the anonymity of respondents can be assured. However, once such indirect techniques are used to measure affective

outcomes in order to certify individual pupils, one can readily predict that students will "learn" to check the expected or acceptable answers. Further, the individual items will be challenged as to their appropriateness for major certification decisions.

If the competencies are to be certified or linked to adult functioning, then the certifying agency may be required to demonstrate that the minimal competency tests have criterion validity. That is, the agency may have to show that the minimal competency tests are actually related to "successful adult functioning."

If the area of employment testing is any indication, then demonstrating criterion validity could prove to be a costly and frustrating task. First, a criterion measure of adult functioning would have to be developed; questions about the validity of this criterion measure would have to be dealt with even before one were to study its relationship to performance on the minimal competency tests. There are any number of ways adults can compensate for some deficit, or lack of skill, and function effectively. The definition of a term like "effectively" is not an easy matter and ultimately is a judgmental value decision. While to my knowledge it has never been required, imagine the problems we would have in showing that the test for a driver's license is related to being a successful driver - whatever that involves! Another problem is that definitions of effective adult functioning can change with technological developments in the society. Old skills can become obsolete, and the need for new skills that we cannot anticipate can arise.

Problems of defining and measuring some criterion of effective adult functioning aside, long-term predictive studies show weak relationships between scholastic or school achievement measures and success in various professions.

Suffice it to say at this point that if the state decides to word its minimal competency testing program in terms of adult functions, then the criterion validity of the certifying test(s) could become an issue. States may be better off avoiding any reference to successful adult functioning when mandating minimal competency programs.

Conclusion

The possible beneficial and negative consequences associated with the institution of a minimal competency testing program need to be anticipated at the outset. Only then can steps be taken to maximize benefits and avoid the very real, detrimental effects that can subtly but inexorably occur with the advent of an external testing program. Since the intent of minimal competency testing is to certify individuals, specific measurement issues, particularly those related to standard setting, scoring, test security and validity need careful attention. For some of these issues there is no easy resolution. However they need to be addressed and resolved as far as possible before a full blown program of minimal competency graduation testing is instituted.

References

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. Taxonomy of educational objectives. N.Y.: McDay, 1956.

Bloom, B.S. & Peters, F.R. The use of academic prediction scales for counseling and selecting college entrants. N.Y.: The Free Press, 1961.

Brimer, M.A., Madaus, G.F., Chapman, B., Kellaghan, T. and Wood, R. Sources of difficulties in school achievement. Windsor Berks, England: NFER Publishing Company, Ltd., 1978.

Commission on Mathematics. Program for College Preparatory Mathematics: Report of the Commission on Mathematics. U.S.A. College Entrance Examination Board, 1959.

Davis, F.B. "School marking procedures." In Educational Measurements and their interpretation. BElmont, California: Wadsworth Publishing Co., Inc., 1964, pp. 287-315.

Edmonds, R. Developing student competency: Alternative means and attendant problems. Harvard University, 1977 (mimeo).

Glass, G.V. Standards and criteria. Paper #10 in Occasional Paper Series. Kalamazoo, Mich.: Evaluation Center, Western Michigan University, 1977.

Greenbaum, W. et al. Measuring educational progress: A study of National Assessment. N.Y.: McGraw-Hill Book Co., 1971.

Holmes, E.G.A. What is and what might be: A study of education in general and elementary in particular. London: Constable & Co., 1911.

Ireland: Department of Education. Report of the Council on Education. Dublin: Stationery Office, 1954.

Jackson, R. Developing criterion-referenced tests. Princeton, N.J.: ERIC Clearing House on Tests, Measurement and Evaluation, Educational Testing Service, 1970.

Jencks, C. The wrong answer for schools is: (b) Back to basics. Washington Post, Sunday, Feb. 19, 1978.

Koerner, J.D. Reform in education: England and the United States. N.Y.: Delacorte Press, 1968.

Madaus, G.F. & Macnamara, J. Public examinations: A study of the Irish Leaving Certificate. Dublin, Ireland: Educational Research Centre, St. Patrick's College, 1970.

McClung, M.S. Competency testing: Potential for discrimination. Clearinghouse Review, August 1977, pp. 439-443.

Morris, G.C. Educational objectives of higher secondary school science. Unpublished Ph.D. thesis, University of Sidney, nd.

Mukerji, S.N. History of education in India: Modern period. Baroda: Acharya Book Depot, 1966.

Norwood Report, 1943. Curriculum Examinations in Secondary Schools. Dublin: Stationery Office, 1943.

Spaulding, F.T. High school and life. N.Y.: McGraw-Hill, 1938.

Srinivasan, J.T. Annual Terminal Examinations in the Jesuit High Schools of Madras, India. Unpublished Ph.D. thesis, Boston College, 1971.

Tyler, R.T. The impact of external testing programs. In W. G. Findley (Ed.), The impact and improvement of school testing programs. Chicago: National Society for the Study of Education, 1963, p. 193-210.